DOCUMENT RESUME

ED 236 827 EC 160 845

AUTHOR Hayden, David L.
TITLE Program Evaluation: Resource Manual.
INSTITUTION Maryland State Dept. of Education, Baltimore. Office of Special Education.
PUB DATE Dec 80
NOTE 192p.; For Appendix H see ED 219 442.
PUB TYPE Guides - Non-Classroom Use (055) -- Information Analyses (070)

EDRS PRICE MF01/PC08 Plus Postage.
DESCRIPTORS *Disabilities; Educational Trends; Elementary Secondary Education; *Evaluation Methods; *Models; *Program Evaluation; *Special Education
IDENTIFIERS Maryland

ABSTRACT

Intended for administrators and evaluators, the manual identifies models useful in evaluating special education programs in Maryland. An introduction to program evaluation, defines the concept of educational program evaluation, notes its purpose, and addresses its current status in the field of special education. Chapter 2 goes into greater depth in examining current trends in program evaluation, reviewing both question-oriented studies (including accountability studies and management information systems) and values-oriented studies (such as policy and consumer-oriented studies). Four comtemporary evaluation models (the Context, Input, Process, Product Model; the Discrepancy Model; the Center for Evaluation Model; and the Countenance Model) are reviewed. Chapter 4 presents a model for special education program evaluation which has four stages: (1) input evaluation of the written program plan, (2) evaluation processes in program implementation, (3) evaluation of products, and (4) analysis of program costs. The final chapter provides guidelines for developing a product evaluation design according to nine steps, from defining the evaluation's goals and objectives to developing the evaluation report. (CL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED236827

PROGRAM EVALUATION

RESOURCE MANUAL

MARYLAND STATE DEPARTMENT OF EDUCATION
DIVISION OF SPECIAL EDUCATION

DAVID W. HORNBECK
State Superintendent of Schools

MARTHA J. IRVIN
Assistant State Superintendent

Principal Author

DAVID L. HAYDEN, Ph.D.
Chief, Information Management

EC160845

DECEMBER, 1980

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
GA Crenson
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

MARYLAND STATE BOARD OF EDUCATION

Members of the Board		Year Term Ends
William G. Sykes President	Baltimore	1981
Lawrence A. Miller Vice President	Baltimore	1982
G. George Asaki	Pasadena	1982
May B. Bolt	Frostburg	1984
Mary Elizabeth Ellis	Salisbury	1983
Verna M. Fletcher	Silver Spring	1984
Joanne T. Goldsmith	District Heights	1985
Albertine T. Lancaster	California	1981
Frederick K. Schoenbrodt	Ellicott City	1985

SECRETARY-TREASURER OF THE BOARD
AND STATE SUPERINTENDENT OF SCHOOLS

David W. Hornbeck

DEPUTY STATE SUPERINTENDENT OF SCHOOLS

Joseph L. Shilling

ASSISTANT STATE SUPERINTENDENT
DIVISION OF SPECIAL EDUCATION

Martha J. Irvin

The Maryland State Department of Education does not discriminate on the basis of race, color, sex, age, national origin, religion, or handicapping condition in matters affecting employment or in providing access to programs. For inquiries related to departmental policy, contact the Office of Equal Opportunity.

Acknowledgements

Recognition and appreciation are extended to the technical consultants, and numerous state and local education personnel who contribu ed to the development of this resource manual and kit.

Dr. Kenneth Olson, Mid-South Regional Resource Center
working in conjunction with the Mid-East Regional
Resource Center

Dr. Charles Greenshield, Auburn University in conjunction
with the Tri-Star Project at the North Carolina Department
of Special Education

Dr. Sang Ahn, Division of Special Education
Maryland State Department of Education

Dr. Stephen Holowenzak, Division of Instruction
Maryland State Department of Education

Miss Judith C. Smith, Division of Special Education
Maryland State Department of Education

Mr. Howard Linaburg, Division of Special Education
Maryland State Department of Education

Ms. Carol Ann Baglin, Division of Special Education
Maryland State Department of Education

Mr. James Harper, Division of Special Education
Maryland State Department of Education

Mrs. Consuelo Anderson, Division of Special Education
Maryland State Department of Education

Mrs. Ruth Reid
Calvert County Board of Education

Dr. Donald Parks
Caroline County Board of Education

Mrs. Mabel Murray
Carroll County Board of Education

Mrs. Lucille Bartenfelder,
Queen Anne's County Board of Education

Mrs. Billie Rinaldi,
Baltimore City Public Schools

Dr. Edward Diefenbach,
Baltimore County Board of Education

PREFACE

Evaluating the progress and future needs of each handicapped child is an essential element in special education. The Congress that passed Public Law 94-142 recognized the fact that intensive effort is needed so that each handicapped youngster may one day acquire the skills to become a functioning member of society. The same legislation also recognized the importance of a school's total special education program by requiring continuous evaluation of its progress.

Obviously, we can observe and evaluate individual children more easily than we can large and complex programs. If we were still living in an age of one-room schools, we would have no particular need for program evaluation. But we are no longer a small society, and we are educating handicapped children to take their place in a world that daily becomes more complex. If we do not evaluate large programs, if we cannot learn and adapt, then we offer no help to children, their parents, or teachers of the future.

Improvement of educational programming is the single most desirable outcome of evaluation. However, program evaluation also assists planning for budgetary, personnel, facility, and materials needs. It can strengthen management's decision-making, it can foster communications, it can clarify objectives. These are the commonsense reasons which helped shape federal legislation.

Since the passage of P.L. 94-142, all state education agencies (SEA) and many local education agencies (LEA) have been developing models for evaluating various parts of special education. Although there are many journal articles and monographs about evaluation theory and model design, there are no guides available that identify some of the procedures for conducting program evaluation that are of particular use for special education.

ODE 5

The major purpose of this resource book is to provide assistance to local special education directors and specialists in conducting evaluation activities which will be compatible and consistent with local and state administrative goals. It is written for the special education administrator, program coordinator, or evaluator responsible for conducting evluations or for using evaluation results in program planning. It will also assist LEA staffs in working more effectively with evaluators. Administrators in many LEAs are not able to have their programs evaluated by professionals, but need to understand the process to make informed decisions about educational programming.

The resource book will first examine definitions of educational program evaluation -- what it is and how it ought to help us plan for the future. It will describe the development of evaluation theory and take a look at current directions in the field. Evaluation models will be examined, and general points on evaluation design will be offered.

The crux of program evaluation in Maryland is found in Chapters 4 and 5. These deal specifically with how to design and develop evaluations for special education programs which can be used by Maryland's school systems. The best current models and philosophies are useless to us if they will not work for special education programming or if they are not adaptable to our federal and state compliance requirements.

We have been continuously engaging in evaluation activities by legal mandate, even though there has been little research and development in this field by special education professionals since the implementation of P.L. 94-142. This book and others like it from other states are going to be the signposts that point to workable special education evaluation methods, useful to local educators and administrators and vital to the improvement of educational programming for handicapped children.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS..... iii

PREFACE..... iv

CHAPTER I

INTRODUCTION TO PROGRAM EVALUATION..... 1

Definition of Program Evaluation..... 1

Purpose of Program Evaluation..... 3

Status of Program Evaluation in the LEA..... 7

CHAPTER II

CURRENT TRENDS IN PROGRAM EVALUATION..... 10

Question-Oriented Studies..... 11

Objective-Based Studies..... 11

Accountability Studies..... 15

Experimental Research Studies..... 16

Testing Programs..... 17

Management Information Systems..... 19

Values-Oriented Studies..... 21

Accreditation-Certification Studies..... 21

Policy Studies..... 22

Decision-Oriented Studies..... 23

Consumer-Oriented Studies..... 25

Client Center Studies..... 27

Connoiseur Based Studies..... 28

Summary..... 30

CHAPTER III

CONTEMPORARY EVALUATION MODELS..... 32

CIPP Model..... 33

Discrepancy Model..... 34

Center for Evaluation (CSE) Model..... 37

Countenance Model..... 39

Summary..... 42

CHAPTER IV

MODEL FOR EVALUATION OF SPECIAL EDUCATION PROGRAMS..... 44

Rationale..... 44

Input Evaluation..... 48

Process Evaluation..... 53

Product Evaluation..... 57

Cost Evaluation..... 61

Summary..... 68

CHAPTER V

GUIDELINES FOR DEVELOPING A PRODUCT EVALUATION DESIGN..... 70

Define Goals and Objectives of the Evaluation..... 73

Develop Evaluation Standards..... 74

Define the Population..... 74

Select Samples..... 75

Select Evaluation Design..... 78

Experimental Design..... 78

Quasi-experimental Design..... 80

Select Data Gathering Instruments..... 87

Develop Evaluation Report..... 97

Summary..... 99

FIGURES

FIGURE 1	Maryland State Department of Education Special Education Program Areas	12
FIGURE 2	Matrix of Types of Studies by Special Education Area	13
FIGURE 3	CIPP Model	33
FIGURE 4	Discrepancy Model	36
FIGURE 5	Stake Model	40
FIGURE 6	Evaluation Model for Special Education Program	47
FIGURE 7	Input Evaluation	52
FIGURE 8	Process Evaluation	56
FIGURE 9	Product Evaluation	60
FIGURE 10	Cost Evaluation	65
FIGURE 11	Flow Chart of Evaluation Procedures	67
FIGURE 12	NASDE Evaluation Design	69
FIGURE 13	Randomized Control Group Pretest and Post-test Design	79
FIGURE 14	One Shot Case Study Design	81
FIGURE 15	One Group Pretest Post-test Design	82
FIGURE 16	Non equivalent Control Group Design	83
FIGURE 17	Pretest Post-test Control Group Design	84
FIGURE 18	Post-test Only Control Group Design	85
FIGURE 19	Post-test Control Group with Pretest and Post-test Design	86
FIGURE 20	Time Series Design	86

APPENDIX

Appendix A: Resource Manual References..... 103

Appendix B: Glossary of Key Special Education and Evaluation Terms..... 113

Appendix C: Questions Raised by Local Directors of Special Education on Program Evaluation..... 130

Appendix D: Cost Evaluation Techniques used by Maryland State Department of Education..... 147

Appendix E: Selected Annotated Bibliography..... 152

Appendix F: A Description of the Materials Included in the Resource Kit..... 162

Appendix G: Overview of the Status of Monitoring and Evaluation Procedures Used by Maryland State Department of Education..... 167

Appendix H: *Standards for Evaluating Educational Programs..... 176

* Appendix H is available as ED 219 442, under the title, "Standards for Evaluations of Educational Programs, Projects, and Materials."

CHAPTER I

INTRODUCTION TO PROGRAM EVALUATION

Definition of Program Evaluation

Exactly what is educational program evaluation? There is no one accepted definition, much less one that fits evaluation of special education programs. Since it is a multi-defined term, it should come as no surprise that its function is also disputed. Should it quantitatively measure, should it focus on performance or on planning? As we look at the development of program planning, we can see the tug between seemingly antithetical approaches. How can scientific measurements be reconciled with the subjective conclusions that professionals must render? Is the best program the one that works and stays within budget? And, if this is so, is to hold to that pragmatic approach useful for long-range planning?

Evaluation is not a new concept to education. In the 1930's, Ralph Tyler demonstrated an approach that laid the foundation for evaluation as we conceive of it today. He viewed evaluation as the process of determining the degree to which the goals of a program have been achieved. The demand for evaluators with a diverse range of expertise has been increasing rapidly ever since. Although there are varied journal articles and monographs about evaluation theory and model design and a variety of texts about psychometrics and statistics, there are still no guides available that effectively synthesize the procedures for conducting special education evaluations.

Although there are still a few educators who subscribe to the measurement definition, an examination of the literature and a review of the various models and classification schemes indicate that model builders and evaluation writers (according to Stufflebeam in 1979) tend to cluster

around two major approaches: (1) those who define evaluation as answering specific questions, such as by assessment of the discrepancy between objectives or standards and an actual performance; and (2) those who define evaluation as the process of obtaining and providing information about the worth of program for decision-makers and consumers. This resource manual will view program evaluation as an analysis of discrepancies between program and objectives or standards and actual performance. Discrepancy evaluation is a broad term which refers to the search for differences between what should be (the standard) and what is (actual performance) [Anderson, Ball, and Murphy, 1975; NASDE, 1977].

The Maryland State Department of Education, Division of Special Education, views evaluation as an analysis of discrepancies between standards or objectives and actual performance. The discrepancies are used as "need indicators" and are analyzed to assist in determining actual needs. The conceptual framework for developing statewide Evaluation systems in special education is the use of status information, such as local or statewide reports, obtained at agreed upon points in time and submitted to decision-makers at state and federal levels to facilitate actions.

This conceptual framework agrees with M.N. Provus (1969) that "the purpose of program evaluation is to determine whether to improve, maintain, or terminate a program. Evaluation is the process of (a) agreeing upon standards, (b) determining whether a discrepancy exists between some aspect of the program and the standards governing that aspect of the program, and (c) using discrepancy information to identify the weaknesses of the program." The generally accepted purpose of evaluation in this context is to systematically identify, collect, analyze, and report information necessary to assess whether an educatioral program should be revised, deleted, left unchanged, or supplemented with additional activities.

The definition we are adopting here must also work within the confines of statutory requirements. The Education for All Handicapped Children Act, P.L. 94-142, delineates the major requirements and general procedures for the evaluation of programs funded under this act. The law requires that each state education agency shall: (a) undertake monitoring and evaluation activities to insure compliance (with the specific requirements of the law) and (b) develop procedures (including specific timelines) for monitoring and evaluating public agencies involved in the education of handicapped children. These procedures must include:

1. Collection of data and reports;
2. Conduction of onsite visits;
3. Audit of federal funds utilization; and
4. Comparison of a sampling of individualized education programs with the programs actually provided. (P.L. 94-142 Regulations, 121a.601, p.42501).

In addition, Maryland law requires the MSDE to monitor all special education programs. These requirements are found in Bylaw 13.04.01, "Programs for Handicapped Children," which states:

> The State Superintendent of Schools shall be responsible for the development of policies and procedures for monitoring and evaluation activities which insure compliance of all public agencies within the state with the requirements of these regulations and State and Federal Legislation. These procedures shall also provide for the evaluation of the effectiveness of the individualized education programs. (13.04.01.03E)

Purpose of Program Evaluation

These mandates require at least two functions: monitoring program implementation and evaluation of program effectiveness. In response to the first requirement, a systematic and comprehensive monitoring and evaluation system was developed in 1978 by the Maryland State Department of Education. The purposes of the monitoring and evaluation process are:

1. To determine the degree of local and state compliance with the policies, standards, and guidelines set forth in the local and state program plans;
2. To determine the areas of either recommended or required change in the programs of special education and related services provided to handicapped children within local education agencies (LEA) and state operated programs (SOP) and to provide a basis for both technical assistance and personnel development;
3. To provide performance data to be used in the ongoing planning and quality improvement of special education programs in the 24 LEAs and SOPs; and
4. To determine the service delivery systems which provide the most effective services at the least cost.

The Monitoring and Evaluation of implementation procedures cycle is through its third phase, and the Division of Special Education is now attempting to more adequately address the second requirement: evaluation of program effectiveness. In addition to Maryland statute 13.04.01.03 E, P.L. 94-142 requires the following:

> Each annual program plan must include procedures for evaluation at least annually of the effectiveness of programs in meeting the educational needs of handicapped children, including evaluation of individualized education programs (P.L. 94-142 regulations, 121a.146, p. 42483).

In addition to the federal procedures already mentioned, Section 618 (B)(2) of P.L. 94-142 specifies that an Office of Special Education shall:

1. Develop effective methods and procedures for evaluation;
2. Test and validate such evaluation methods and procedures; and
3. Conduct actual evaluation studies designed to test the effectiveness of such programs and projects.

These requirements are obviously difficult to meet. In effect, they ask us to justify the state's philosophical approach to evaluation and the methodology we have selected. There has been little research in special education program evaluation since the passage of P.L. 94-142, as mentioned before. Special education programs have not been routinely evaluated. Models continue to be set forth, yet within the mandates of funding constraints, questions remain about how an administrator is to design a program evaluation system and how to set realistic criteria for conducting special education program evaluation. Olsen (1979) discusses the problems that all states are grappling with as they work toward compliance in "Evaluating Effectiveness of Special Education Programs".

It is now time to look at our current evaluation efforts to judge their effectiveness and plan for future modifications. Beyond simple implementation compliance, our objective should be a communication system that allows local administrators to provide the most effective educational programs for handicapped children. Evaluation should be a tool to inform LEAs that state standards are met, their goals and objectives are being appropriately implemented, and the associated costs are within budget constraints. To insure the usefulness of evaluation results, LEA evaluators should base designs on a thorough understanding of program purposes and local organization. More specifically, special education administrators can use program evaluation information:

- to meet documentation and reporting requirements of state and federal regulatory agencies;
- to understand specifically what a program is intended to accomplish, what methods are useful, and what results are being produced;

- to determine whether a program's objectives are attainable and relevant or need modification; and
- to decide whether the resources available for a program are adequate for achieving the results expected.

A well-designed and well-conducted evaluation provides information that improves the process of decision-making. It eliminates or at least greatly reduces the influence of political or self-serving factors, and it provides objective, defensible evidence. Evaluation can lead to the planning of more effective programs when data-based evidence of what is working and what is not working is available to program planners. Evaluation increases the likelihood that decisions will be wise and that subsequent policy will be rational.

Evaluation is not without its problems. Every method contains its own strengths and weaknesses. A high level of precision is possible when one quantitatively counts but is diminished when one is required to make subjective judgments. Evaluation can be costly requiring a great deal of time and skill.

Evaluators should not be surprised if, when administrators are not actively involved in planning specific evaluation studies, reports are laid neatly on a shelf or are buried altogether when results are negative or run counter to vested interests. Few administrators are readily willing to accept evaluation results that may place the survival of their program in serious jeopardy. Only when organizational personnel themselves are dissatisfied with a program and have viable alternate programs will they be receptive to the implications of a negative evaluation and take its results seriously.

Evaluators would be wise to use simple public relations techniques that increase the potential usefulness of an evaluation:

1. Identify potential users early in the evaluation and address issues of concern to them;
2. Involve representatives from different constituencies in the process of evaluation;
3. Complete the evaluation according to schedule;
4. Prepare several forms of the report, including a nontechnical summary for general audiences;
5. Provide individuals whose program is being evaluated with a draft report so they will have an opportunity to critique the report and provide relevant supplementary data;
6. Take responsibility for presenting the findings to decision-makers and translating them into possible action plans; and
7. Be available for advice or assistance in implementing the recommendations even after the evaluation has been completed. (Rose & Nyke, 1977)

Throughout education, innovative projects and programs are being implemented in classrooms, schools, LEAs, and state agencies. Some will work very well, some will be a waste of everyone's time and limited funds. There is no way of knowing what works unless the time is taken to check. Evaluation can provide specific information about a program. Knowing what results a program is producing is one of the best bases for making effective decisions.

Status of Program Evaluation in the LEA

In Baglin's (1979) survey of LEAs within Maryland, there appears to be a major trend toward developing and/or adapting program evaluation models from regular education and making them specific to special education needs.

Efforts have been underway within all LEAs to evaluate each of the P.L. 94-142 discretionary projects through the required application format. In fact, several LEAs indicated in this survey that the only programs that are being evaluated now are the federal programs. Most LEAs are attempting to systematically evaluate inservice programs for staff development through the use of MSDE standardized questionnaires for the Comprehensive System of Personnel Development projects.

There are several LEAs who are making major efforts toward the development of a Data Base Information Evaluation System for special education. In 1978, Montgomery County installed a computer based student tracking system, the Continuum Education Data System. This system has a discrepancy model design and was developed to provide information to assist management with decision-making and fulfilling accountability responsibilities.

The Continuum Education Data System satisfies several purposes. It is being used to generate reports documenting the delivery of services in relation to needs for handicapped and economically disadvantaged students. These reports are also used by the Office of Continuum Education to analyze current services and to plan and budget for future needs. Directors, program supervisors, principals, and others are provided with specific information about students with special needs as well as services recommended and provided for those students.

In 1976, Anne Arundel County installed its own computer-based student tracking system to provide individual and summative data on types of services being recommended and provided to each handicapped student. This discrepancy model tracking system collects the same data as the MSDE's Special Services Information System (SSIS). This system is used by administrators to assist in program management, and provides data needed for LEA planning and allocations of staff and resources.

Baltimore County has a project in cooperation with the Maryland State Department of Education to develop an evaluation design to provide information about the effectiveness of special education programs in terms of student performance on standardized achievement tests. Each student's achievement test results are analyzed relative to other students of the same age. Two special education levels of service (IV, and V) are involved. This evaluation of program effectiveness includes the monitoring of pupil progress as measured by norm-referenced group ability and group achievement tests. The Metropolitan Achievement Test, 1978 edition, and the Otis Lennon School Ability Test (OLSAT) were selected to provide measures of student achievement and ability, respectively. Fifteen hundred students, ages 7 to 13 years, have been involved in the evaluation study. There are two major goals for the project. First, the data will be used by teachers to monitor each pupil's progress over a period of years. Second, the data will be aggregated to facilitate program evaluation. In addition, it is hoped that this design can serve as a model for program evaluation for specific curriculum areas throughout the state.

The next chapter will describe where current thinking on evaluation is leading us. It includes some prototypes being developed and some general guidelines for successful evaluation design.

Chapter II

CURRENT TRENDS IN PROGRAM EVALUATION

The field of educational evaluation has developed dramatically over the past ten years. Following a period of relative inactivity in the 1950's, educational evaluation efforts experienced revitalization in the middle 1960's, influenced by articles of Lee Cronbach (1963), Daniel Stufflebeam (1966), Malcolm Provus (1971), Michael Scriven (1967), and Robert Stake (1967). The field's development was further stimulated by the evaluation requirements of the "great society" programs that were launched in 1965, by the nationwide accountability movement that began in the early 1970's, and most importantly, by the mounting federal requirements for evaluation information relative to federally funded projects.

Of particular interest for this review are the diverse ways in which educational evaluation has been conceptualized and the different approaches by which educational institutions have operationalized evaluation. The different theoretical approaches to evaluation are characterized and assessed, and different patterns of evaluation practice have emerged in various school systems during the past decade.

Stufflebeam (1979) classified these eleven types of studies into two main approaches. The first approach includes evaluations that are oriented to answer specific questions whose answers may or may not assess a program's worth. The second approach involves studies that are designed primarily to assess and/or improve the worth of some program.

These two main approaches include eleven types of studies: five are of the questions-oriented approach, and six are of the values-oriented approach. He has further analyzed each study in terms of descriptors: the

main purpose served, the source of questions, the questions characteristic of each study type, and the methods typically used.

The special education program areas were developed based on the analysis of P.L. 94-142 and Maryland Bylaw 13.04.01 using a program planning and budget system. The twenty-one program areas listed in Figure 1 have been developed for use by LEAs in drawing up their annual comprehensive special education plans and by MSDE for monitoring and evaluating program implementation.

Figure 2 presents a matrix which identifies the various types of evaluation studies that may be suitable for evaluating each of the special education program areas and examples of current articles reflecting each type of evaluation. The matrix also identifies the individual who has pioneered the concept and the individuals who have been involved in furthering the development of the specific type of study.

QUESTIONS-ORIENTED STUDIES

Questions-oriented studies reflect the MSDE, Division of Special Education's major role in program evaluation. Questions-oriented studies are those which start from a particular question to be addressed (due to requirements of federal and state laws), move to the methodology appropriate for answering that question, and only then consider whether the question and methodology are appropriate for developing and supporting value claims. In some cases, their focus is too narrow or is only tangential to a program's worth.

Objectives-Based Studies

In the objectives-based study, some statement of objectives provides the advance organizer.* The objectives may be ones that have been mandated

*Advance Organizer - A term used by systems analysts which indicates the framework within which a study takes place.

Figure 1

MARYLAND STATE DEPARTMENT OF EDUCATION
SPECIAL EDUCATION PROGRAM AREAS

	Program Area	Legal Citations
1.	Child Identification	P.L. 94-142 Regulations (Section 121a.220)
2.	Confidentiality	P.L. 94-142 Regulations (Section 121a.560-121a.576) Annual Program Plan (Part VII F)
3.	Full Educational Opportunities	P.L. 94-142 Regulations (Section 121a.123)
4.	Facilities, Personnel, & Service	P.L. 94-142 Regulations (Section 121a.223)
5.	Personnel Development	P.L. 94-142 Regulations (Section 121a.380-121a.383)
6.	Parent Involvement	P.L. 94-142 ReguLations (Section 121a.226)
7.	Participation in Regular Education (LRE) (Least Restrictive Environment)	P.L. 94-142 Regulations (Section 121a.550)
8.	Public Control of Funds	P.L. 94-142 Regulations (Section 121a.228)
9.	Excess Cost	P.L. 94-142 Regulations (Section 121a. 182-121a.186) Article of the Annotated Code of Maryland (Section 106-106G)
10.	Nonsupplanting	P.L. 94-142 Regulations (Section 121a.230)
11.	Comparable Services	P.L. 94-142 Regulations (Section 121a.231)
12.	Information - Reports	P.L. 94-142 Regulations (Section 121a.232-121a.233)
13.	Records	P.L. 94-142 Regulations (Section 121a,233)
14.	Public Participation	P.L. 94-142 Regulations (Section 121a.226, 245, 560-573 & 500-508)
15.	Individualized Education Program	P.L. 94-142 Regulations (Section 121a.340-121a.346)
16.	Policies Consistent with Statute	P.L. 94-142 Regulations (Section 121a.236)
17.	Nondiscrimination	P.L. 94-142 Regulations (Section 121a.239)
18.	Procedural Safeguards	P.L. 94-142 Regulations (Section 121a.50 subpart B)
19.	Nonpublic	P.L. 94-142 Regulations (Section 121a.140 and subpart D)
20.	Voc.-Ed. Career Education	ESSA Act, Part B, Section 613 (a)(2)
21.	Priorities	P.L. 94-142 Regulations (Section 121a.127)

A MATRIX FOR TYPE OF STUDY BY SPECIAL EDUCATION PROGRAM AREAS

*Indicates that a program evaluation may be applicable to the type of study.

Special Education Program Areas \ Type of Study	Value Oriented						Question Oriented				
	Accreditation	Policy Study	Decision Oriented	Consumer Oriented	Client Centered	Connoisseur Based	Objective Based St.	Accountability	Experimental Studies	Testing Program	MIS
1. Child Identification	*	*	*	*	*	*	*	*	*	*	*
2. Confidentiality		*		*	*			*			*
3. Full Educational Opportunities		*	*	*	*	*	*				*
4. Facilities	*	*	*	*	*						*
5. Personnel Development	*	*	*	*		*	*	*		*	*
6. Parent Involvement		*	*	*	*		*				*
7. Participation in Regular Class (LPE)		*	*	*	*	*	*	*	*	*	*
8. Public Control of Funds		*	*	*			*	*			*
9. Excess Cost		*						*			*
10. Nonsupplainting		*	*	*	*			*			*
11. Comparable Services	*	*	*	*	*		*	*			*
12. Information/Report	*	*	*				*				*
13. Records		*	*					*			*
14. Public Participation		*	*	*			*	*			*
15. Individualized Education	*	*	*	*	*	*	*	*	*	*	*
16. Policy consistant with Statute		*					*	*			*
17. Nondiscrimination	*	*	*	*	*		*	*			*
18. Procedural Safeguard	*	*	*	*	*		*	*	*		*
19. Nonpublic	*	*	*	*			*	*			*
20. Vocational Career Ed.	*	*	*		*	*	*		*	*	*
21. Priorities		*	*	*	*		*	*			*
Procedures Developers	Cooperative Study of secondary school standard (1933)	Coleman Jenks Clarks Owens Wolf	Alkin, Ashburn Brickell, Estes Guba Merriman, Otto Reinhard	Glass	MacDonald Rippy Guba	Guba Sanders	Bloom, Hammond Metfessel & Michael, Popham Provus	Stennerd Webster Kearney	Suchman Wiley, Glass	Flanagan Novick Hyronymus Thorndike	Kauffman
Concept Pioneer	College Entrance Examination Board (1901)	Rice	Cronbach Stufflebeam	Scriven	Stake	Eisner	Tyler	Lessinger	Lindquist Cambell & Stanley Cronbach	Lindquist	Cook Keer
Current Special Education Articles Reflecting Type of Study	Anderson	Holloway	Faraley, MacMillan, Howe, Turner Blackhurst	Education Advocates, ACC	Grotsky, Adelman, Hauser, Jones	BEH	White, Meyer, Gallagher, Olsen, Smith Holovenzak	Olsen	Dunst	Salvia	Borich Lilly Meierhenry, NASDE

13

23

2-1

by the client, ones that have been developed by the evaluator, or ones that have been developed by teachers. An example of this type of evaluation is the P.L. 94-142 requirement that every handicapped student receiving special education must have an annual written individualized educational program (IEP) listing objectives and criteria for evaluating each objective. Also, each P.L. 94-142 discretionary project must have written measurable objectives. The purpose of an objectives-based study is to determine whether the objectives have been achieved. It is usually program developers, sponsors, and managers who want to know the answers, and to paraphrase Ralph Tyler's characterization of the National Assessment Program (NAP), the clients of objectives-based studies usually want to know which students have achieved which educational objectives.

The methods used in objectives-based studies essentially involve the collection and analysis of performance data relative to specified operational objectives.

The objectives-based type of study has been the most prevalent type used in the name of educational evaluation. It is one that has good commonsense appeal, educators have had a great amount of experience with it, and it makes use of technologies of behavioral objectives and standardized testing. Common criticisms are that the approach leeds to terminal information that is of little use in improving a program and that this information often is too narrow in scope to constitute a sound basis for judging the merit of an educational program.

Some current special education articles reflecting this type of study are White (1978), Meyer (1972), Gallagher (1973), Holowenzak (1978), Olsen (1979), and Smith (1980).

- Holowenzak, Hayden, and Harper describes a matrix sampling method for doing input and process evaluation for special education program areas 1 and 15 (Child Identification and Individual Education Programs) using an objectives based type of evaluation study.

- Olsen describes methods of doing input, process, and product/effectiveness evaluation for special education programs using objective-based type of evaluation study.

- Smith describes methods of doing product evaluation for special education program area 5 (Personnel Development) using a question-oriented approach with an objectives based type of evaluation study.

- White describes a method for doing product evaluation for special education program area 15 (Individual Education Program) using a question-oriented approach with an objectives based type of study.

- Gallagher describes a method of doing input, process, and product evaluation for preschool age special education programs using an objectives based type of study.

Accountability Studies

The accountability study became prominent in the early 1970's. Its emergence might have been connected with widespread disenchantment with the lack of significant results from the increasing stream of evaluation reports. One approach to rectifying the situation requires that accountability systems could insure that educators would carry out their reponsibilities to improve education and that evaluators would do a thorough job of identifying all present effects of their programs.

The advance organizer for the accountability type of study is a written set of responsibilities that education personnel are assigned for implementing educational programs. The purpose of the study is to provide constituents with an accurate accounting that the procedures were performed according to the written plan and to account for products identified in the plan.

The questions that are addressed in accountability studies come from the constituents of programs, such as taxpayers, parent groups, school boards, and local, state, and national funding agencies. The main question the groups want answered concerns whether the involved personnel and organizations charged with responsiblity for educating students and for

improving education are achieving all they should be, given the investment of resources to support their work. Methods used in an accountability study include procedures for auditing the design, process, results of evaluation studies, mandated testing programs, and performance contracting.

Some examples within special education are reviewing (auditing) LEAs' Special Education Manuals, IEP forms design, public control of funds, etc.

The main advantages of the accountability-type of study are that it is very popular among constituent groups and is aimed at improving the quality of education. Main disadvantages are that it produces a good deal of political unrest and acrimony among professional educators and their constituents. Also, political forces may tend to force the implementation of accountability efforts before the technology needed to carry them out can be developed, field tested, and put into action on a sound basis.

Current special education articles reflecting this type of study are Olsen and Linde (1980), and Vaughan, (1975).

- Olsen and Linde describe a method of doing product evaluation for secondary special education programs using an accountability type of study.
- Vaughn describes methods of doing cost evaluation of products for special education programs using an accountability type of study.

Experimental Research Studies

The experimental research type of study is quite prominent in the education evaluation. The advance organziers in experimental studies are problem statements, hypotheses, and investigatory questions. The usual purpose of the experimental research study is to determine causal relationships between a specified independent and dependent variable, such as a given instructional method and student standardized test performance.

It is particularly noteworthy that the source of questions investigated in the empiracal research study are asked by researchers and

developers of educational programs and not usually the constituents, practitioners, or financial sponsors. Persons who pioneered in using experimental types of studies are E.F. Lindquist (1953), Donald Campbell and Julian Stanley (1963), and Lee Cronbach and Richard Snow (1969).

An example of experimental question-oriented studies being conducted by the Maryland State Department of Education, Division of Special Education is:

> Do different types of handicapped students receiving the same special education level of programming perform differently on the Maryland Functional Reading Test?

The main advantage of using an experimental study in evaluation work is that it provides strong methods for establishing causal relationships between treatment and student achievement variables. Problems, however, are that the approach usually is not workable in field settings and provides a much narrower range of information than is needed to evaluate entire education programs. In addition, the approach tends to provide terminal information that is not useful for guiding the future development of the program.

A current special education article reflecting the type of study is Dunst, (1979).

- Dunst describes methods for doing <u>product evaluation</u> for three special education program areas 1, 3, and 15 (Child Identification, Full Educational Opportunity, and IEP) using an experimental-design type of evaluation study.

<u>Testing Programs</u>

Since the 1930's, American education has been inundated with standardized testing programs. Probably every school district in the United States has some such testing program, and formerly, many educators tended to equate the results of a standardized testing program with the

information needed to evaluate the quality of a school district, a school, a program, and in some cases, even a teacher.

Advance organizers for the standardized testing program type of study include areas of the school curriculum, tests that are available from publishing companies, and specified norm groups. The main purpose of testing programs is to compare the test performance of an individual student or a group of students to that of selected norm groups.

The source of questions to be answered is usually test publishers and test selection committees. The typical question addressed by tests concerns whether the test performance of individual students is at or above the average performance of local, state, and national norm groups. The main process involved in using testing programs is to select, administer, score, and report standardized test results.

The main advantages of standardized testing programs for students who are similar to the norm population are that they are efficient in producing valid and reliable information of student performance in many areas of the school curriculum and that they have been used at every level of the school program in virtually all districts in the United States.

The main limitations are that they provide data only about student performance and that they reinforce students' multiple-choice test-taking behavior rather than their writing and speaking behavior. For specific types of handicapped students, there are unique concerns that tests must accommodate the handicapping condition. Also, many tests do not have special norms for handicapped students so that there is no way of knowing whether the instruments are validated for specific handicapped groups.

A current special education book dealing with the use of norm tests with handicapped students is Salvia and Ysseldyke (1978).

- Salvia and Ysseldyke describes methods for doing input and product evaluation for special education program areas 1 and 15 (Child Identification and Individual Education Programs) using a testing type of study. The authors review in detail underlying constructs, and the strength and weakness of various types of tests when used with handicapped students.

Management-Information Systems

The management-information system is the final type of questions-oriented study. It is similar to an accountability study, except that it supplies managers with the information they need to conduct programs based on the Program Implementation Plan as opposed to supplying them with the information they need only to win a political advantage or requirement of law. The management-information type of study is also like the decision-oriented study which will be discussed later, except that the decision-oriented study provides information both to develop and to define the merit of a program.

The advance organizers in most management-information systems include program objectives, specified activities, and projected program milestones or events. The purpose of a management-information system, as already implied, is to supply continuously the information managers need to plan, direct, and control their programs.

The source of questions addressed by a management-information system is management personnel. The main question they typically want answered is: "Are program activities being implemented according to schedule, and budget, and with the expected results?"

Methods commonly used in management-information systems are systems analysis, Program Evaluation and Review Technique (PERT), Critical Path Methods (CPM), Program Planning and Budgeting Systems (PPBS), Management by Objectives (MBO), computer-based information systems, and cost analysis.

Desmond Cook (1966) introduced the use of PERT in education, and Roger Kaufman (1969) has written widely about the use of management information

systems in education. Holowenzak (1976) describes how many of these types of techniques are being used in Maryland. MSDE, DSE has made commitments to this type of system over the past ten years with its SSIS and the use of PPBS format for developing the LEA Comprehensive Plan.

A major advantage of the use of management-information systems is in giving managers information they can use to plan, monitor, and control complex operations. A major difficulty with the application of this private industry type of system to education is that the products of education do not conform to a narrow, precise definition as is the case with a corporate profit and loss statement. Even when the information gathered in management-information systems lacks the scope required to assess worth, it provides one of the best means for managing complex programs like special education with its many state and federal requirements such as developing and evaluating the IEP for each handicapped student.

Some current special education articles reflecting this type of evaluation study are Borich (1977), Lilly (1978), Meierhenry (1969), and NASDE (1978).

- Borich describes methods for doing input evaluation of the results of the planning, designing, and developing phase of special education programs using a Management Information System type of evaluation.
- The National Association of State Directors of Special Education Report on ten local education agencies that describes methods of doing input, process, and product evaluation of special education programs using information-management system type of evaluation study with computers.
- Lilly describes a method of doing process evaluation for the special education program area 15 (Individual Education Program) using a Management-Information System type of evaluation study.
- Meierhenry describes methods of doing input, process, and cost evaluation for special education programs using a Management Information System type of evaluation study.

VALUES-ORIENTED STUDIES

Accreditation-Certification Studies

Most educational institutions have periodically been the subject of an accreditation study, and most professional educators have had, at one time or another, to meet certification requirements for a given educational position. Such studies of institutions and personnel are in the realm of values-oriented evaluation efforts, since institutions and personnel are studied to prove whether they are fit to serve designated functions in society.

The advance organizers used in the accreditation-certification type of study usually are guidelines that have been adopted by some accrediting or certification body. As previously suggested, the purpose of the studies is to determine whether institutions, programs, or personnel should be approved to perform specified functions.

The source of questions for accreditation or certification studies is the accrediting or certifying agency. Basically, the questions they ask are: "Are institutions programs, and personnel meeting minimum standards, and how can their performance be improved?"

Typical methods used in the accreditation-certification type of study are self-study and self-reporting by the individual or institution. In the case of institutions, panels of experts are assigned to visit the institution, verify a self-report, and gather additional information. The basis for the self-studies and the visits by expert panels are usually guidelines that have been specified by the accrediting agency.

Accreditation of education was pioneered by the College Entrance Examination Board in about 1901.

Examples within special education of this type of study are state approval of nonpublic schools for providing education programs to

handicapped students, approval of staff certification to work in special education, and approval of construction of Level V special education facilities.

The main advantage of the accreditation or certification type study is that it aids lay persons in making informed judgments about the quality of educational institutions and the qualifications of education personnel. A main difficulty is that the guidelines for accrediting and certifying bodies typically emphasize the process of education and not the product of education. Also the self-study and visitation process used in accreditation offers many opportunities for corruption and/or misinterpretation of performance data.

An article reflecting this type of study is Anderson and Sonnen (1973).

- Anderson and Sonnen describes a section of the National Study of Secondary School Evaluation criteria for the evaluation of secondary schools for doing input and process evaluation of special education programs which uses an accreditation type of evaluation study.

Policy Studies

The policy study is the second type of values-oriented evaluation. It is recognized as such because it sets out to identify and assess for society or some segment of society the merits of competing policies.

The advance organizer for the policy type of study is a policy issue, such as, "What is the best way to meet federal guidelines for equal educational opportunity or non-biased assessment of handicapped students?" The purpose of the policy study is usually to identify and assess the potential costs and benefits of competing policies for a given institution or for society.

Legislators, policy boards, and special interest groups often posit the questions that are addressed by policy studies. A main question they

may pose is, "Which of two or more competing policies will maximize the achievement of valued outcomes at a reasonable cost?"

Methods that are often used in policy studies include the Delphi Technique (1960), experimental and quasi-experimental design (as in New Jersey's 1974 negative income tax experiment), scenarios, forecasting, and judicial proceedings.

An example within special education of this type of study is the Maryland Formula Review Study for determining the excess cost of special education in each LEA.

Joseph Rice can be mentioned as the pioneer in the area, as he conducted massive studies around 1900 to help education decide on the merits of continued concentration on spelling in a school district curriculum.

The main advantage of policy studies is that they provide information directly related to decision-making needs of institutions and society. The main problem is that policy studies are often corrupted or subverted by the political environment in which they must be conducted and reported.

A current special educaton article reflecting this type of study is Holloway (1980).

- Holloway examines the method of doing input and process evaluation of special educator's policy analysis of the special education area 7 (Participation in Regular Class/LRE) using a values oriented approach incorporating a policy-evaluation type of study.

Decision-Oriented Studies

The decision-oriented study emphasizes that evaluation should be used actively to help improve the worth of a program as well as retroactively to judge its worth. As mentioned previously, the decision-oriented study should be distinguished from management-information systems because of the former's emphasis on questions of worth.

Decision situations provide the advance organizer for decision-oriented studies. Basically, the purpose of the studies is to provide knowledge and a value base for making and defending decisions.

The sources of questions addressed by these studies are involved decision-makers, including administrators, parents, students, teachers, school boards, taxpayers, and all others who make decisions about funding, conducting, and using the results of education. Many times the Office of Special Education (federal) uses public forums to obtain more information and recommendations from various advocate groups.

Many methods may be used in a decision-oriented study. The two acceptable methods for obtaining information include needs assessments and case studies. An example within special education is the Admission, Review and Dismissal Team review of the IEP.

Lee Cronbach (1963) first introduced educators to the idea that evaluation should be reoriented from its objectives-based history to a concern for helping educators make better decisions about how to educate.

A main advantage of the decision-oriented approach is that it encourages educators to use evaluation continuously and systematically in their efforts to plan and implement programs that meet educational needs. It also presents a rationale for helping educators to be accountable for decisions they have made in the course of implementing a program.

A main limitation is that the collaboration required between an evaluator and decision-maker introduces opportunities for biasing the evaluation results. However, countermeasures, including external review, can be introduced to offset tendencies toward bias.

Some current special education articles reflecting this type of evaluation study are Parsley (1971), MacMillan (1977), Blackhurst (1979), and Turner (1978).

- Blackhurst describes a method of doing product evaluation for special education program area 5 (Personnel Development) using a values oriented approach which incorporates a decision-oriented type of evaluation study.

- MacMillan describes methods for doing input, process and product evaluation for special education program area 7 (Participation in Regular Classes/LRE) using a values-oriented approach with a decision-oriented type of evaluation study.

- Parsley describes a method for doing an input, process and product evaluation for special education programs using a decision-oriented type of evaluation study.

- Turner describes a method for doing process and product evaluation for the special education program area 15 (Individual Education Program) using a decision-oriented type of evaluation study.

Consumer-Oriented Studies

In the consumer-oriented approach, the evaluator is the "enlightened surrogate consumer." Advance organizers are societal values and needs. The purpose of a consumer study is to judge the relative merits of alternative educational goods and services, thereby helping taxpayers and practitioner groups make wise choices in their purchases.

Questions for the consumer-oriented type of study are derived from society at large, from constituents of educational institutions, and especially from the evaluator's frame of reference. The general question addressed is: "Which of several alternative consumable education projects is the best buy, given costs, the needs of the consumer group, and the values of society at large?" Typical methods include goal-free evaluation, quasi-experimental designs, modus operandi analysis, and cost analysis. A popular method is having an independent consumer advocate conduct the evaluation and report the findings on studies of publicly supported educational programs.

Michael Scriven (1967) pioneered the consumer-oriented approach in education, and there are strong parallels between his work and the concurrent work of Ralph Nader in the general field of consumerism. Gene Glass (1969) has been an avid supporter and developer of Scriven's work.

Examples of this type of study in special education are the Education Advocates Coalition Report (1980) on Implementation of P.L. 94-142 and American Coalition of Citizens with Disabilities (1979) study on vocational training for the handicapped.

- American Coalition of Citizen with Disabilities (1979) Report is an example of a method of doing a needs assessment as input evaluation of the special education program areas 7 and 2 (Particpation in Regular Education, and Vocational Education) using a value-oriented approach with a consumer-oriented type of evaluation study.

- The Education Advocates Coalition (1980) Report is an example of a method of doing process evaluation of the special education program areas 1, 2, 6, 7, 15, and 18 (Child Identification, Confidentially, Parent Involvement, Participation in Regular Education (LRE), IEP, and Procedural Safeguards) using a value-oriented approach with a consumer-oriented type of evaluation study.

A main advantage of the approach is that it is a hard-hitting, independent assessment intended to protect educators from shoddy educational products and services. The approach has high credibility with consumer groups.

The main disadvantage of the approach is that it can be so independent from practitioners that it may not assist them to do a better job of serving consumers. Also, the consumer-oriented type of study requires a highly credible and competent expert plus sufficient resources to allow the expert to conduct a thorough study. Often it is too costly to be carried out well, and thus produces faulty, unrealistic data.

Client-Centered Studies

In direct contrast to consumer-oriented studies is the client-centered study. The client-centered study takes the local autonomy view and helps people who are involved in a program to evaluate it and use that evaluation in planning program improvement.

The advance organizers used in the study are concerns and issues in the program itself. The purpose of the study is to help people in a local setting understand the operations of their program and the ways they are valued by people who are experts in the program.

Typical methods used in the client-centered study are the case study, adversary report, socio-drama, and what Robert Stake has called "responsive evaluation" (1970).

Stake is the pioneer of the client-centered study, and his approach has been further developed by Barry McDonald (1975), Robert Rippy (1973), and, most recently, Egon Guba (1978).

An example within special education is the self-evaluation conducted by a LEA using MSDE monitoring and evaluation procedures on an off year of the three year state site visit cycle. Another example is the P.L. 94-142 Discretionary Projects Evaluation Component completed by the LEA.

The main strength of the approach is that it is an action-research approach in which people implementing programs are helped to conduct their own evaluation. Its main weaknesses are its lack of external credibility and its susceptibility to bias on the part of people in the local setting, since they, in effect, have great control over the evaluation study.

Some current special education articles reflecting this type of evalution study are Grotsky (1974), Adelman (1970), Hauser (1979), and Jones (1978).

- Adelman describes a method of doing process evaluation of special education program area 5 (Personnel Development) using a value-oriented approach incorporating a client-centered type of evaluation study.

- Grotsky describes a method of doing continuous process evaluation for special education programs using a client-centered type of evaluation study. This Peer Evaluation Program uses teachers and support staff input for self evaluations and provides criteria for use in both self-and peer-evaluations.

- Hauser describes a method of doing process evaluation for special education program area 7 (Participation in Regular Classes) using a values-oriented approach with a client-centered type of evaluation study.

- Jones describes methods of doing product and cost evaluation for the special education program area 7 (Participation in Regular Classes) using a values-oriented approach with a client-centered type of study of attitudes, a questions-oriented approach with an accountability type of study concerning cost evaluation, and a testing-type of evalution study concerning the achievements of students.

Connoisseur-Based Studies

The connoisseur-based study is the most recent entry in the field of study types. It is based upon the assumption that certain experts in a given field are capable of indepth analysis and evaluation that could not be done through other approaches. Just as a national survey of wine drinkers would undoubtedly yield information concerning wine preferences, it would not provide the detailed judgment of quality that might be derived from a single connoisseur who has devoted a professional lifetime to wines.

The advance organizer of the connoisseur-based type of study is the evaluator's special expertise and sensitivities. The purpose of the study is to critically describe, appraise, and illuminate the particular merits of a given object. The methodology of the approach includes the systematic use by evaluators of their perfect perceptual sensitivities, past experiences, and refined insights. Judgment is then conveyed to help the audience appreciate and understand all of the nuances of the object under study.

The main advantage of the connoisseur-based study type is that it exploits the particular expertise and finely-developed insights of persons who have devoted much time and effort to the study of the precise area. They can provide an array of detailed information that the audience can use to form a more insightful analysis than otherwise might be possible. The disadvantage of the approach is that it is solely dependent upon the expertise and qualifications of the particular expert doing the evaluation, leaving much room for subjectivity, bias, and corruption. There is a danger, too, that the connoisseur will not gear the language of the final report to the client's understanding.

Some current special education articles reflecting this type of study are the four monographs by the Office of Special Education (1978).

- The Office of Special Education conducted a needs assessment study developing criteria for input evaluation dealing with special education program areas 1, 7, 15 and 18 (Child Identification, Least Restrictive Environment, IEP, and Procedural Safeguards) using a values-oriented approach incorporating decision-oriented and connoisseur-based types of evaluation studies.

SUMMARY

No type of evaluation study should be considered inherently bad. Education occurs in a political world, and authorities responsible for its programs must be politically astute enough to operate within the boundaries of public opinion. Further, studies will be public relations-oriented and stress the program's best features. In addition, studies that excel in addressing given questions, whether or not those questions constitute a comprehensive assessment, are desperately needed. Each of the type of studies described here can play a constructive role in education. Features of one might be combined with another to enhance particular advantages.

The main problem with the questions-oriented studies is that the answers collected are more narrow than the information needed for a values-oriented assessment of worth. However, these studies frequently are superior to values-oriented evaluation studies in the efficiency of methodology and technical adequacy of information employed.

Values- oriented studies have much to recommend them, since they are a comprehensive attempt to assess worth. However, because they undertake such an ambitious task, they are, in many respects, doomed to fall short of expectations due to the impossibility of assessing something as subjective as true worth.

In studying advantages and disadvantages of these various types, evaluators may encounter considerable difficulties if they and their clients have dissimilar perceptions of the study being undertaken. Typically, clients may want a politically advantageous, values-oriented study performed, while the evaluator may prefer the questions-oriented approach in which he has had professional training. If evaluators are

ignorant of the likely conflict between perceived goals, the evaluation is probably doomed to failure from the start. An evaluator should advise all involved parties of possible conflicts over a study's methodology and objectives and should negotiate a common understanding at the start. Finally, it is imperative to remember that no one type of study is consistently the best approach to evaluating education.

Chapter III

CONTEMPORARY EVALUATION MODELS

This section provides a definition of model, some characteristics of a good model, dangers in the use of models, and brief descriptions of four of the best known educational evaluation models influencing the Division of Special Education, the Maryland State Department of Education.

Van Dalen defines models as "simplified or familiar structures which are used to gain insights into phenomena that scientists want to explain." (1966) Models may be drawings or physical replicas that represent the real things, or they may be more abstract. Mathematical equations, verbal statements, symbolic descriptions, graphic presentations, or electromechanical devices may be used to represent objects and relationships that are being modeled. Some investigators locate a structure about which much is known and use it to gain insight into a field about which little is known.

The term model is also defined as "a structure that enables one to present concepts in such a way that researchers/evaluators can gain useful insight/information into their phenomena." The model includes:

1. Structural elements (input, process, product),
2. Program flows (from planning, through implementing, to ending), and
3. Criteria/standards for program evaluation.

Some characteristics of good models are that they describe and help us to understand complex systems and provide the framework within which experiments and evaluations are done. Models help us to see new relationships as well as help to predict when experiments will be impossible.

Some dangers in the use of models are that they invite over-generalization, and entice us into committing logical fallacies. The relationships between variables may be incorrect or not validated. Model building often diverts useful energy into non-productive activity.

THE CIPP MODEL

One of the best known of the decision-facilitation evaluation schemes is the CIPP Model (Context, Input, Process, Product). The CIPP model was developed by Daniel Stufflebeam and Egon Guba (1971).

The CIPP defines evaluation as the process of "delineating, obtaining, and providing useful information for judging decision alternatives." Stufflebeam identifies three steps involved in his model: delineating and providing are carried out collaboratively between evaluator and decision-maker, while obtaining information is a technical activity carried out primarily by the evaluator.

The model relates to four decision settings and evaluation types:

1. Planning to determine objectives-context evaluation;
2. Structuring to design instructional procedures-input evaluation;
3. Implementing to use, monitor, and improve the procedures-process evaluation; and
4. Recycling to judge and react to the outcomes produced by the procedures-product evaluation.

Figure 3

CIPP MODEL

	Context	Input	Process	Product
Decision Making	Proactive (Formative) Evaluation			
Accountability	Retroactive (Summative) Evaluation			

THE DISCREPANCY MODEL

Malcolm Provus, (1971) during his tenure as Director of Research for the Pittsburgh, Pennsylvania public schools, devised another model for a systematic approach to evaluation based on the premise that evaluation involves the comparison of performance with standards. More specifically, Provus (1971) defines evaluation as:

1. Defining program standards;
2. Determining whether a discrepancy exists between some aspect of the program and standards governing that aspect of the program; and
3. Using discrepancy information either to change performance or program standards.

The Discrepancy Model identifies five major stages of program development and three major content categories for a program evaluation.

The five development stages can be described as dealing with (a) definition, (b) installation, (c) process, (d) product, and (e) cost. The process of evaluation consists of moving through developmental stages and content categories in such a way as to facilitate a comparison of program performance with standards while at the same time identifying standards to be used for future comparisons. The three content categories are (a) input, (b) process, and (c) output.

At Stage 1, - Definition - the evaluation task is to obtain a definition of the program based on the program content taxonomy. The definition obtained becomes the program performance information to be compared with the content taxonomy. A discrepancy between any component in the program definition and the same component in the taxonomy represents evaluation information to be used by those responsible for the nature and effectiveness of the program.

At Stage 2, - Installation - the standard for comparison is the program definition arrived at in Stage 1. Program performance information consists of observations from the field regarding the installation of the program's components. Discrepancy information is used to redefine the program or to change installation procedures.

At Stage 3, - Process - the standard is that part of the program definition which describes the relationship between program processes and enabling objectives. Discrepancy information is used either to redefine process and relationship of process to interim product or to better control the process being used in the field.

At Stage 4, - Product - the standard is that part of the program definition which refers to terminal objectives. Program performance information consists of criterion measures used to estimate the terminal effect of the project.

The last stage, Stage 5, - Cost Analysis - represents all cost benefit options available to the evaluation only after the first four stages have been negotiated. Notice that the use of discrepancy information always leads to a decision to:

1. Go on to the next stage,
2. Recycle the stage after there has been a change in the programs standards or operation,
3. Recycle to the first stage, or
4. Terminate the project.

The generalizable content of a program is shown in Figure 4: the transactions that transform input into output. Equally important are precise estimates of the amount of time and money needed to locate and use resources.

In all stages of the development of evaluation of educational programs, the content specifications remain the same: input, process, and output relative to time and money. Taken in their form in Figure 4, they represent a taxonomy.

Figure 4

DISCREPANCY MODEL

TAXONOMY OF PROGRAM CONTENTS

Content Categories	Time	Cost
Input: Staff qualification by position Staff preprogram training StudEnt selection criteria Student - entry behavior Media Facilities Administrative Condition	Acquisition Time:	Budgeted Cost:
Process: Student transactions with students, staff, media facilities Staff transactions with staff, students, media, facilities, administration, others	Operating Time:	Operating Costs:
Outputs: Enabling Objectives (EO) Terminal Objectives (TO) Ultimate Objectives (UO) Interrelationships between EO, TO, UO	Cumulative Time	Cumulative Costs

THE CENTER FOR THE STUDY OF EVALUATION (CSE) MODEL

The UCLA Center for the Study of Evaluation (CSE) is a research and development center devoted to the analysis and improvement of educational evaluation practice (Alkin 1969). The CSE model is geared to the requirements of those who must make educational decisions, as is indicated in their definition of evaluation:

> Evaluation is the process of determining the kinds of decisions that have to be made, selecting, collecting, and analyzing the information needed in making these decisions; and reporting this information to appropriate decision-makers. (Alkin 1969)

The CSE Model is similar to the CIPP Model except that what Stufflebeam refers to as "process evaluation" has been substantially altered. Whereas the CIPP evaluator focuses on procedural considerations during the process evaluation, the CSE Model encourages evaluators to attend to the enroute products as well as the process of the program being evaluated. The CSE approach is more explicitly tied to the different kinds of decisions to be made at each of five stages.

The CSE Model identifies five stages of evaluations:

1. Needs assessments
2. Program planning
3. Implementation evaluation
4. Progress evaluation
5. Outcome evaluation

Needs Assessment

The initial stage focuses on provision of information regarding the extent to which educational programs are meeting their objectives. An attempt is made to ascertain the current status of an educational program's outputs. Then this status (the performance of learners) is contrasted with desired program outcomes. The program needs are the difference between what exists and what is desired. The desired stages are used to identify program goals. The decision focus of this first stage of the CSE Model is problem selection.

Program Planning

The second stage of the CSE Model provides information regarding the sorts of instructional programs that meet the needs identified during the needs assessment stage. This is done by identifying resources, strategies, and timelines for achieving specific objectives.

Implementation Evaluation

After program planning, information is gathered on the degree to which the instructional program is actually being carried out in accordance with the program plan. This stage of the CSE is most akin to CIPP's process evaluation. The decision focus of the CSE Model is program modification.

Progress Evaluation

The fourth stage of the CSE Model provides information on the extent to which the planned program is achieving its objectives. The evaluator tries to identify the relative success of the different components of a program with a view to improving deficiencies. Unlike the CIPP, however, CSE evaluators also examine products of the program at this stage. The decision focus of progress evaluation is also program modification.

Outcome Evaluation

The final stage of the CSE Model emphasizes the provision of information regarding the general worth of the program as reflected by the outcome it produces. Information regarding the extent to which objectives have been achieved is obviously significant. The decision focus of the final stage of the CSE Model is program certification/adaptation.

A comparison of the CSE and the CIPP Models reveals several similarities, particularly with respect to the first two stages of each model. There are also important differences that will become even more pronounced as Stufflebeam and his associates sharpen the CIPP Model and Alkin and his colleAgues refine the CSE Model. Information regarding materials descriptive of the CSE approach can be secured from the Center for the Study of Evaluation. (Alkin, 1969)

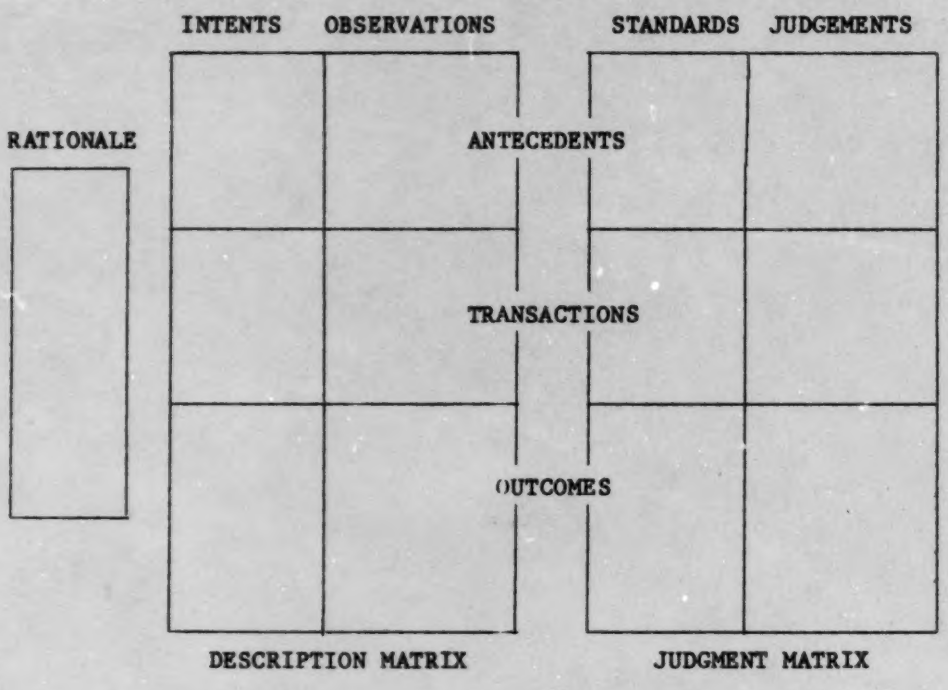
THE COUNTENANCE MODEL

Stake's conception of evaluation (1967) emphasizes two chief operations: description and judgment.

One of the helpful aspects of the model is that it distinguishes between descriptive and judgmental acts of the evaluator according to three phases of an educational program: its antecedent, transaction, and outcome phases. The antecedent is conditions existing prior to instruction that may relate to outcome. Stake views transactions as the "succession of engagement" that constitute the process of instruction. Outcome is considered to be the effect of an instructional program.

Figure 5

Stakes depiction of the statements and data needed by evaluators



Stake then divides descriptive acts according to whether they refer to what was intended or what was actually observed. He argues that both intentions and what actually took place must be fully described. He then divides judgmental acts according to whether they refer to the standards used in reaching judgments or to the actual judgments themselves. He assumes the existence of some kind of rationale guiding the design of an educational program.

As seen in Figure 5, Stake combines these several distinctions in a graphic representation of the statements and data needed to be gathered by an evaluator. Notice that, apart from the rationale, the single cells at the left of the layout refer to descriptive operations, whereas the single cells at the right deal with judgmental operations.

Stake's data matrices help us to see how his model inclines the evaluator to engage continually in description and judgment. His distinction between intents and observations reminds us how the aspirations of educators are often not realized in the antecedents, transactions, and outcomes that actually occur.

After devising his 1967 model, Stake became more attuned to discovering what the evaluator's clients actually want. He recommends a responsive evaluation stance in which great attention is given to the description and judgment needs of those who commission the conduct of an evaluation. Stake believes that insensitivity to the perceived needs of those individuals for whom an evaluation is being conducted will lead to an unresponsive, futile evaluation.

SUMMARY

The final stage of each model is concerned with a cost benefit evaluation of the product of the education program with other competing programs. Provus properly points out that unless there is a sufficient degree of consistency in the product associated with the programs to be compared, such cost-benefit evaluations are impossible. For example, comparing a neurological evaluation with a vision evaluation instead of two like evaluations (products).

As can be seen from a comparison of the Discrepancy Model with the three other models (CIPP, CSE, Countenance) previously examined, there is much similarity among these four evaluation frameworks. They all subscribe to the same general conception of evaluation's function - to provide information to those making decisions about programs. These four models conceive of the educational evaluation as providing information needed by the decision maker.

Special Education Administrators should always refer to the National Program Evaluation Standards developed by NIE (National Institute of Education), when reviewing or designing an evaluation study using one of the above models. These standards are summarized as follows:

Accuracy Standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features of the object being studied that determine its merit.

The Utility Standards are intended to ensure that an evaluation will serve the practical information needs of given audiences.

The Propriety Standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

The Feasibility Standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

Models such as the four described serve as a framework for an evaluation program. Maryland has been developing its evaluation model using various components of these described models to meet specific reporting requirements.

Chapter IV

MODEL FOR EVALUATION OF SPECIAL EDUCATION PROGRAMS

Rationale:

The need for the development of an evaluation model for special education programs emanates from the Maryland State Bylaw 13.04.01.

The Bylaw specifies state responsibility for the evaluation of special education programs as follows:

> The State Superintendent of Schools shall be responsible for the development of policies and procedures for monitoring and evaluation activities which insure compliance of all public agencies within the state with the requirements of these regulations and state and federal legislation.

This statement indicates that Maryland must develop evaluation standards for program evaluation and procedures in which the evaluation action can take place. Evaluation standards are criteria for judging attainment of the already established program goals and objectives in the program.

The special education program areas were developed based on the analysis of P.L. 94-142 and Maryland Bylaw 13.04.01 using a program planning and budget system. The twenty-one program areas were developed for use by LEAs in drawing up their annual comprehensive special education plans and by MSDE for monitoring and evaluating program implementation.

Each MSDE program area has one or more objectives listed which are required by law. Each LEA must specify the objective measure/standard for each objective. The LEA may select any number of activities and resources to implement the particular objective.

An evaluation model is needed in order that administrators can gain useful insight/information about special education. The MSDE evaluation model identifies three stages of program status (input, process, and product) and generates discrepancy information by comparing the exisitng program status with the standards according to federal and state mandates. The fourth stage of the MSDE model deals with a cost analysis, which is also required by federal and state law. Cost analysis at the input stage is budget evaluation; at the process stage it is quarterly expenditure reports; and at the product stage it is the final cost analysis report.

The conceptual framework for developing A statewide evaluation model in special education is the use of status information obtained at agreed upon points in time which become state reports submitted to decision-makers, at the state and federal levels, to facilitate actions.

The MSDE evaluation model focuses on information needed by these decision makers to answer the following questions:

1. Are programs planned in accordance with federal and state regulations?
2. Are programs being implemented in accordance with the plan?
3. Are the intended beneficiaries being served?
4. In what settings are the beneficiaries being served?
5. What services are being provided to beneficiaries?
6. Has the program brought or attained the expected results?
7. What are the consequences of implementing federal and state requirements?
8. What administrative procedures are in place?
9. Has the program brought reasonable results based on its comparable cost efficiency indices?

DESCRIPTION OF THE EVALUATION MODEL

The evaluation model for special education programs consists of four evaluation stages. This model is formulated with concepts and procedures from the current evaluation models of the CIPP Model (Stufflebeam, 1974), the Discrepancy Model (Provus, 1971), the CSE Model (Alkin, 1969), (Popham, 1975) and many others.

The first stage of the model deals with the program input evaluation of the written program plan (needs assessment, planning, and budgeting data); the second stage of the model deals with the evaluation of processes in implementing the program; the third stage evaluates products of the programs; and the fourth stage analyzes cost of the program (See Figure 6).

Procedures for the evaluation must be stated in terms of specific actions to be performed. The state defines the purposes of the monitoring and evaluation as a procedure:

1. to determine the degree of local and state compliance with the policies, standards, and guidelines set forth in the local and state program plans;

2. to determine the areas of necessary change in handicappeD children's services and resources within the 24 LEAs and provision of technical programmatic assistance as well as personnel development;

3. to provide a basis for ongoing planning and quality improvement of special education programs in the 24 LEAs;

4. to determine the service delivery systems which provide the most effective services at the least cost; and

5. to determine the extent to which our programs are cost efficient and cost effective.

Figure 6

EVALUATION MODEL FOR SPECIAL EDUCATION PROGRAM

	Stage 1 Input	Stage 2 Process	Stage 3 Product	Stage 4 Cost
Purposes	1. Accreditation 2. Approval 3. Certification	1. Monitoring 2. Readjustment of Ongoing Program	1. Program Effectiveness	1. Cost Efficiency 2. Cost effectiveness
Methods	1. Define Program (federal, state regulations) compare to the program status (Documents) 2. Needs Assessment	1. Assessment of timelines and procedures 2. Measure participant's perceptions of program and implementation steps	1. Criterion Reference Test and standardized achievement test 2. Attitudes or perception inventories	1. Compare unit cost with program (cross sectional) 2. Compare unit cost yearly basis. (longitudinal)
Specific Evaluation	1. Goal Analysis 2. Policy Analysis 3. Accreditation Evaluation 4. Needs Assessment for a New Program	1. Attitude, perception toward the program 2. Assessment of implementation of procedures and timelines	1. Degree of goal attainment or standards met 2. Degree of behavioral changes in attitudes and perception.	1. Cost efficiency cross sectional longitudinal 2. Cost effectiveness comparision of quantitative outcomes: cross sectional, and longitudinal

Stage 1. Input evaluation

The input of the program is defined as the identification of educational resources and planned programs. It will include:

Program Plans
Goals and Objectives
Personnel/Staff/Teacher
Students
Parents
Budget
Media
Facilities
Transportation
Administrative Conditions

All of these elements interact to help attain predetermined goals under certain environmental conditions. Input evaluation is applicable to current MSDE review of special education programs for approval, designed to answer the question, "Does the program have the required resources?" Provus (1969) lists the following elements of an input evaluation:

1. Staff qualifications by position,
2. Staff pre-program training,
3. Student selection criteria,
4. Student entry behavior,
5. Media/Materials,
6. Facilities, and
7. Administrative conditions.

P.L. 94-142 regulations, Section 121a.402, require that SEAs be involved in program approval activities for private schools. In determining whether or not schools should be state approved, MSDE staff members review such things as qualifications of staff and teachers, types of handicapped children served, facilities, materials, equipment, and the administrative procedures available for operating a particular school. A

site team visits the private school to review the program in relation to documents submitted. Stake (1967) draws a parallel between input evaluation and program accreditation. Most accrediting teams use a process similar to that described above for special education private school program approval. In Maryland, LEA comprehensive plans and policies are approved when there is agreement with state and federal laws and regulations. Therefore, the approval process is primarily an input evaluation to determine whether or not a program has the prerequisites for adequate service delivery. Another example of input evaluation at the state level is approval of new special education facilities or Level V centers.

Maryland procedures require that the LEA conduct a needs assessment of the type, number, location, and educational requirements of Level V students. This LEA Needs Assessment Report must be reviewed and verified by the MSDE Division of Special Education as part of standard procedures set by the MSDE Inter-Agency Committee Public School Construction Program (IAC). Once the need for a Level V center has been established, the design, location, and budget must be included in the LEA's construction plan in order that the IAC may complete its input evaluation stage.

An example of input evaluation at the LEA level is the IEP approval for each handicapped student needing special education. In determining whether a student's IEP should be approved, the ARD team members review such things as the qualified examiner's report documenting specific types of handicapping conditions, educational assessment reports documenting special education needs, IEP's goals, objectives, staff requirements, timelime for implementation, and the location of the special education services. The admission function of the ARD Team deals with input

evaluation, the review function of the ARD Team deals with process evaluation, and the dismissal functional of the ARD Team deals with product evaluation.

The input evaluation stage for each of the MSDE 21 program areas is designed to answer the following questions:

1. Is the program defined with a written document?

   A program must have a written document which defines that program's goals and objectives, staff, students or target population, facilities, instruments, resources, schedules, and the terminal objectives to be achieved. This program definition is then used as the standard for the process evaluation.

2. Is the program proposal designed in compliance with federal and state policies and regulations?

   State education agencies must evaluate the program proposal with state requirements.

3. Are the program definitions operationalized?

   The definition of the program must be stated in terms of measurable and tangible terms of quantity or quality.

4. Is the program definition consistently identified with the real needs of target populations?

   The needs assessment of the input evaluation is concerned with determining whether the input is consistent

with the needs of the target population (students, parents, or staff) as defined by state and federal law. Any discrepancies will indicate an inadequacy.

5. Is the budget adequate to purchase needed resources?

As seen in Figure 7, depending on the questions to be answered and the target population, the procedure will differ as to what is relevant. When starting an input evaluation for a particular program area, the first step is to obtain a written program definition, which lists goals and objectives, target population, necessary resources, administrative procedures, etc.

The second step is to determine consistency with federal/state standards and real needs of the target population, as in state evaluation for accreditation of private schools and LEA Comprehensive Plan approval activities.

Figure 7

INPUT EVALUATION

Figure 7

Program Areas	Question 1	Question 2	Question 3	Question 4	Question 5
	Is the program proposal defined with written document?	Is the program proposal designed in compliance with federal and state regulations?	Are definitions operationalized?	Is the program definitions consistently identified with the real needs of the target population?	Is the budget adequate to purchase needed resources?
1. Child Identification					
2. Confidentiality					
3. Full Educational Opportunity					
4. Facilities, Personnel & Service					
5. Personnel Development					
6. Parent Involvement					
7. Participation in Regular Education					
8. Public Control of Funds					
9. Excess Cost					
10. Nonsupplanting					
11. Comparable Services					
12. Information - Reports					
13. Records					
14. Public Participation					
15. Individualized Education Program					
16. Policies Consistent with Statute					
17. Nondiscrimination					
18. Procedural Safeguards					
19. Nonpublic					
20. Voc.-Ed. Career Education					
21. Priorities					

Stage 2. Process evaluation

The purpose of the process evaluation is to determine whether the program was installed as outlined in the plan and delivers the services as defined by its objectives. The evaluation is conducted by reviewing program implementation documents and by asking participants (teachers, students, administrators, and parents) their perceptions of program objectives and operating procedures.

Olsen (1979) suggests that process evaluation for special education is what is mainly being done through SEA monitoring activities. Monitoring is defined by the National Association of State Directors of Special Education (NASDSE) as:

> A continuous review procedure designed to compare present functioning against specified standards. The procedure should yield a profile showing areas of conformance as well as discrepancies requiring new procedures, training, or other methods where improvement is needed in order to comply with specific standards.

Olsen lists the following series of elements that should be part of the process evaluation of special education programs.

1. Student transactions
   a. Students
   b. Staff
   c. Instructional and assessement materials, equipment, and supplies
   d. Facilities
   e. Transportation

2. Staff transactions
   a. Staff
   b. Students
   c. Parents
   d. Instructional and assessment materials, equipment, and supplies
   e. Facilities
   f. Administration
   g. Other

3. Child-centered transactions
   a. Child identification, location, and evaluation
   b. IEP development
   c. Educational placement
   d. IEP implementation
   e. Evaluation of child performance and IEP review

4. Program support transactions
   a. Procedural safeguards
   b. Service coordination and communications
   c. Instructional and assessment material delivery systems
   d. Personnel training and development
   e. On-going planning, policy development, management and evaluation

Examples of the first kind of transaction evaluated at the MSDE level through the SSIS tracking of handicapped students are: (1) number of hours interacting with nonhandicapped students (LRE Report), (2) number of hours interacting with other handicapped students in special education (OCR Report), (3) number of interactions with special education staff and related services personnel, and (4) number of interactions with Vocational Education and Vocational Rehabilitation personnel, (5) number of students in Level V and VI (special centers and residential facilities).

An example within special education of the second kind of transaction in process evaluation monitored at the state level is staff case load listing the number of handicapped students served. (SSIS)

Examples of the third kind of transactions monitored by the state through the Individual Case Review (ICR) and SSIS process are: (1) number of new students identified and evaluated (SSIS Child Find Report), (2) number of students served in private and parochial schools, and (3) number of IEPs developed and implemented.

An example of the fourth kind of transaction monitored by the state is the number of staff participating in inservice training activities (CSPD).

In implementing the monitoring (or process evaluation) component of program evaluation, MSDE personnel determines standards from laws, policies, procedures, and attorney general opinions from state and federal agencies. It then constructs a series of questions which interprets the standards. The monitoring and evaluation team reviews plans and policies of the LEA and compares these plans and operating procedures with state and federal laws and regulations. In monitoring actual service delivery, the M & E team uses the logic that the approval of plans is always provisional based on a school or LEA providing the services that it said it would provide.

Using the above assumptions, the process evaluation for special education is defined as identifying the discrepancy between the existing program status and the expected standard. Measurement is done through the participants' perception of program implementation, such as procedure, schedules, or the program itself.

As seen in Figure 8, According to the questions to be answered and the subject/target population, procedures will differ as to what is relevant.

When starting a process evaluation for a particular program area, the first step is to review the written documents concerning program goals and objectives, its target population, implementation time lines, approved budget allocation, existing quarterly reports, and any special projects impacting upon the program area. The five basic questions to be answered for each program area are:

1. Is the program delivering the same number and kind of services according to the design?
2. Are services delivered within time lines?
3. Are procedures practical and working to participants' satisfaction?
4. Are there systematic data control procedures.
5. Are budgeted resources being expended according to approved plan?

Figure 8

PROCESS EVALUATION

Figure 8

Program Areas	Question 1 Is the program delivering the same number and kind of services according to the design?	Question 2 Are services delivered within timeline?	Question 3 Are procedures practical and working to participants satisfaction?	Question 4 Are there systematic data control procedures?	Question 5 Are budgeted resources being expended according to approved plan?
1. Child Identification					
2. Confidentiality					
3. Full Educational Opportunity					
4. Facilities, Personnel & Service					
5. Personnel Development					
6. Parent Involvement					
7. Participation in Regular Education					
8. Public Control of Funds					
9. Excess Cost					
10. Nonsupplanting					
11. Comparable Services					
12. Information - Reports					
13. Records					
14. Public Participation					
15. Individualized Education Program					
16. Policies Consistent with Statute					
17. Nondiscrimination					
18. Procedural Safeguards					
19. Nonpublic					
20. Voc.-Ed. Career Education					
21. Priorities					

Stage 3. Product Evaluation

Product evaluation has been neither well defined nor distinguished from impact, output, outcome, or effectiveness evaluation. The reason for the lack of distinction is that most writers according to Olsen, 1979, using the terms output, outcome, product, effect, and impact in program evaluation are referring to the same entity. Product evaluation is defined as:

> Changes in persons or organizations which can be measured or described in terms of quantity, quality, or duration and which are attributable to a program. Such changes may occur relative to the student, relative to the satisfaction of persons with the program, or relative to the organization of the program.

The interchangeable use of the terms output, outcome, product, effect, and impact is demonstrated by the definitions used by authorities in the field of evaluation. Anderson, Ball, and Murphy's Encyclopedia of Evaluation (1975) equates evaluation of effectiveness with product evaluation. They indicate that this form of evaluation is used to measure and interpret program attainments usually related to specified goals. This definition of product evaluation is consistent with Stufflebeam (1971, 1974) who relates products to results. Definitions of product evaluation may correspond to those used by other writers using the terms outcome and output evaluation (Alkin, 1969; Gephart; Stake, 1967; Provus, 1969, 1971). Gephart (date not available) has demonstrated that these various writers have actually presented complementary and, in many cases, overlapping concepts rather than contradictory ones. Level (1975) indicates that his cost-effectiveness model simply adds the element of cost of the outcome of a program, and Scriven's Goal-Free Model (1972) for determining side effects adds no new concepts which are not covered by the above definition.

Insufficient attention has been paid by special educators to the need for evaluating the products of special education programs and demonstrating their effectiveness. Gallagher and Hall (1979) indicate that minimum competency test results may serve as one source of data for demonstrating product effectiveness.

Olsen (1979) lists the following series of elements that should be considered as part of the product effectiveness evaluation of special education:

1. Behavior Changes in Children and Changing the Child's Context

    a. Goal Related Change re
        (1) Child's IEP
        (2) Competency test results
        (3) Curriculum or scope and sequence
        (4) High school graduation
        (5) Self-sufficiency

    b. Perceptions of or interactions with child by
        (1) Family
        (2) School staff
        (3) Regular students and lay public

2. Participant Satisfaction

    a. Statement of Satisfaction
        (1) Student
        (2) Teachers
        (3) School administrators and supervisors
        (4) Parents
        (5) Other taxpayers/lay public
        (6) Legislators

    b. Indications of Satisfactions
        (1) Selection as model
        (2) Personnel turnover
        (3) Acceptance of referrals and recommendations
        (4) Involvement by others
        (5) Replication

3. Internal and External System Organizational Changes

    a. Service Availability re
        (1) School policies or funding
        (2) Access to services (educational and support)
        (3) Range of services

    b. Service Facilitation
        (1) Level of communication (within and without)
        (2) Level of cooperation
        (3) Redundancy of services

Some examples of the third type of product evaluation conducted at the state level are staff availability within LEA to meet standards concerning teacher student ratio by levels according to the ByLaw 13.04.01 (SSIS data), availability of specific services to meet individual student needs as recommended by the ARD Committee (SSIS Data), LEA Child Find system by the number of new students identified (SSIS Child Find Data), and the LEA screening and diagnostic evaluation services as compared to time line standards in ByLaw 13.04.01 (ICR Data).

Finally, service availability and service structures at the LEA level could be reviewed and compared with other LEA systems or a previously operated system of its own to determine whether or not programs are resulting in increase in services, increased cooperation, reduction of redundancy, increased communication, more awareness of services, and replication of exemplary programs.

This third stage of the MSDE evaluation model deals with program effects on the target areas (students, parents, staff, and community) by measuring the behaviors of cognitive and affective domains. At this stage, the evaluator may need to use various experimental or quasi-experimental designs for arriving at an answer to the question, "Has the program achieved its terminal objectives?" Chapter V provides a brief description of the eight most commonly used experimental and quasi-experimental designs.

The five basic questions to be answered in each program area for product evaluation are shown in Figure 9.

Figure 9

PRODUCT EVALUATION

Figure 9

Program Areas	Question 1	Question 2	Question 3	Question 4	Question 5
	Has the program effected on the target population changed?	Was the effect consistent enough and observed after thought to be educationally as well as statistically significant?	Were analysis of quantitative and qualitative information completed using standards procedures?	Were results objectively reported and conclusion justified?	Was cost of the product produced by the program clearly stated?
1. Child Identification					
2. Confidentiality					
3. Full Educational Opportunity					
4. Facilities, Personnel & Service					
5. Personnel Development					
6. Parent Involvement					
7. Participation in Regular Education					
8. Public Control of Funds					
9. Excess Cost					
10. Nonsupplanting					
11. Comparable Services					
12. Information - Reports					
13. Records					
14. Public Participation					
15. Individualized Education Program					
16. Policies Consistent with Statute					
17. Nondiscrimination					
18. Procedural Safeguards					
19. Nonpublic					
20. Voc.-Ed. Career Education					
21. Priorities					

Stage 4. Cost Evaluation

Cost evaluation is concerned with all three stages: input/budget stage, process/expenditure stage, and product/cost effective stage, an expenditure and its outcome. A cost analysis (evaluation) for a program involves terminologies such as efficiency, effectiveness, and benefit. Vaughan (1975) defines these terms as follows:

Cost-efficiency is a conceptual framework for analyzing the standard cost of various approaches, services, and methods with the actual costs of each delivery approach for reaching predetermined objectives. The efficiency ratio obtained demonstrates the monetary return on the monetary investment. A formula is

$$\frac{\text{Standard Cost/Product Unit or Output}}{\text{Actual Cost /Product Unit or Output}}$$

Cost-effectiveness is a conceptual framework for analyzing effectiveness of human and non-human resources, programs, services, or delivery modes by comparing, in ratio form, achieved results to predetermined objectives or product. The ratio indicates how various alternative approaches or services compare. A formula for cost effectiveness analysis is

$$\frac{\text{Achieved Output or Product/Time}}{\text{Planned Output or Product/Time}}$$

Cost-benefit is a conceptual framework for analyzing and synthesizing the cost-efficiency and cost-effectiveness ratios to determine the short and long term value and benefits to society for offering any special education programs and services. Thus, cost-benefit incorporates the potential and actual productivity of the handicapped child who participated in special education programs.

Adding cost analysis to an evaluation format has two advantages: it helps service consumers understand how efficiently money is being spent, and helps managers choose the most effective services at a cost within budget.

The use of a PPBS structure for the LEA Comprehensive Plan and for P.L. 94-142 projects facilitates cost evaluation at the input, process and product stages. Cost evaluation at the input stage deals with evaluating program and line item budgets. Cost evaluation at the process stage deals with auditing expenditures according to line items and timelines for using resources. Cost evaluation at the product stage is evaluating the total and unit cost of each product achieved.

Once definitions of program areas, and goals are accepted, work can begin to develop a complete plan for each program area. Administrative personnel in each special education program area will develop a written program plan having (1) program objectives, (2) objective measures, (3) implementation procedures, including necessary activities essential to reaching the program objectives, (4) description of resources - human and non-human - required to operate the programs, and (5) a budget.

Cost-effectiveness evaluation of a product is possible when there is agreement on the type of students with the same learning characteristics and outcome performance/product.

Completing the analysis of program objectives requires appropriate cost data which can be assigned to each special education program. Basically, there are two forms of cost -- direct and indirect. Once elements within each of these two categories are established, data collection and cost allocation are relatively easy.

Direct costs are those elements included in the school LEA which are necessary for the program to operate. They may be limited but will be part of the reimbursable cost of special education.

Indirect costs are those items which are integral components in education and must be prorated. They remain an added expense to the LEA as found in this example from Colorado. (1976)

1.0 benefit pay	4.0 heat, lights, phones
2.0 insurance	5.0 custodial services
3.0 travel	6.0 regular education supervision and administration

See Appendix D for Maryland description of direct & indirect cost.

Line items, whether direct or indirect costs, can be allocated by programs. In some instances, allocation of partial costs is necessary. The objective is to define and allocate all direct/indirect costs by special education program area. When a student's program is funded by 2 or more sources, they must be identified to determine the total or unit cost. Until this is completed, a true fiscal picture cannot be seen.

Outcome or product refer directly to the special education program objectives. Thus, the stated outcome must be defined clearly within the program objectives and tied to objective measures and standards which include, under what conditions, how well, and judged by whom or by what evaluation measure.

If it is being determined that there are differences in the alternative programs being offered, adequate data must be available to describe how each delivery system functions under which constraints.

Cost allocations are assigned to program objectives as a third component in data analysis. It is essential that all direct/indirect costs be prorated, development, or evaluative costs be accumulated, prorated, and properly assigned for determining cost-efficiency. When the prescribed number of cost-effective/cost-efficient ratios are computed, then the data can be analyzed, synthesized into categories, and then supplied to decision-makers.

Cost benefit data, when added to efficiency/effectiveness data should provide a definitive picture of the value of an educational product. Political, social, economic, legal, and technical pressures may influence decision making policies or practices within a given special education program. Therefore, by reliance upon a systems approach for obtaining relevant, timely, and accurate measures of efficiency cost, in addition to the evaluation of product effectiveness, will lend logic to program selection. If special education program are designed to meet the needs of the handicapped students and prepare them for successful entry into a constantly changing community, in-depth analysis or needs assessment the societies' goals and values are mandatory. Any reduction in the probability of making untenable decisions only reinforces the need for a Systems Approach to special education evaluations.

As seen in figure 10 according to the questions to be answered and the subject/target population's procedures will differ as to what is relevent. The four basic questions to be answered for each program area are:

1. Is the program budget cost efficient? (input)
2. Are the funds expended as planned? (process)
3. Are the program products cost effective? (product) and
4. Are the program products cost beneficial? (product)

Figure 10

COST EVALUATION

Figure 10

Program Areas	Question 1 Is the program budget cost efficient? (Input)	Question 2 Are the funds expended as planned? (process)	Question 3 Are the program products cost effective? (product)	Question 4 Are the program products cost beneficial? (product)	
1. Child Identification					
2. Confidentiality					
3. Full Educational Opportunity					
4. Facilities, Personnel & Service					
5. Personnel Development					
6. Parent Involvement					
7. Participation in Regular Education					
8. Public Control of Funds					
9. Excess Cost					
10. Nonsupplanting					
11. Comparable Services					
12. Information - Reports					
13. Records					
14. Public Participation					
15. Individualized Education Program					
16. Policies Consistent with Statute					
17. Nondiscrimination					
18. Procedural Safeguards					
19. Nonpublic					
20. Voc.-Ed. Career Education					
21. Priorities					

Program evaluation at the state level can be viewed as four stages as seen in Figure 11.

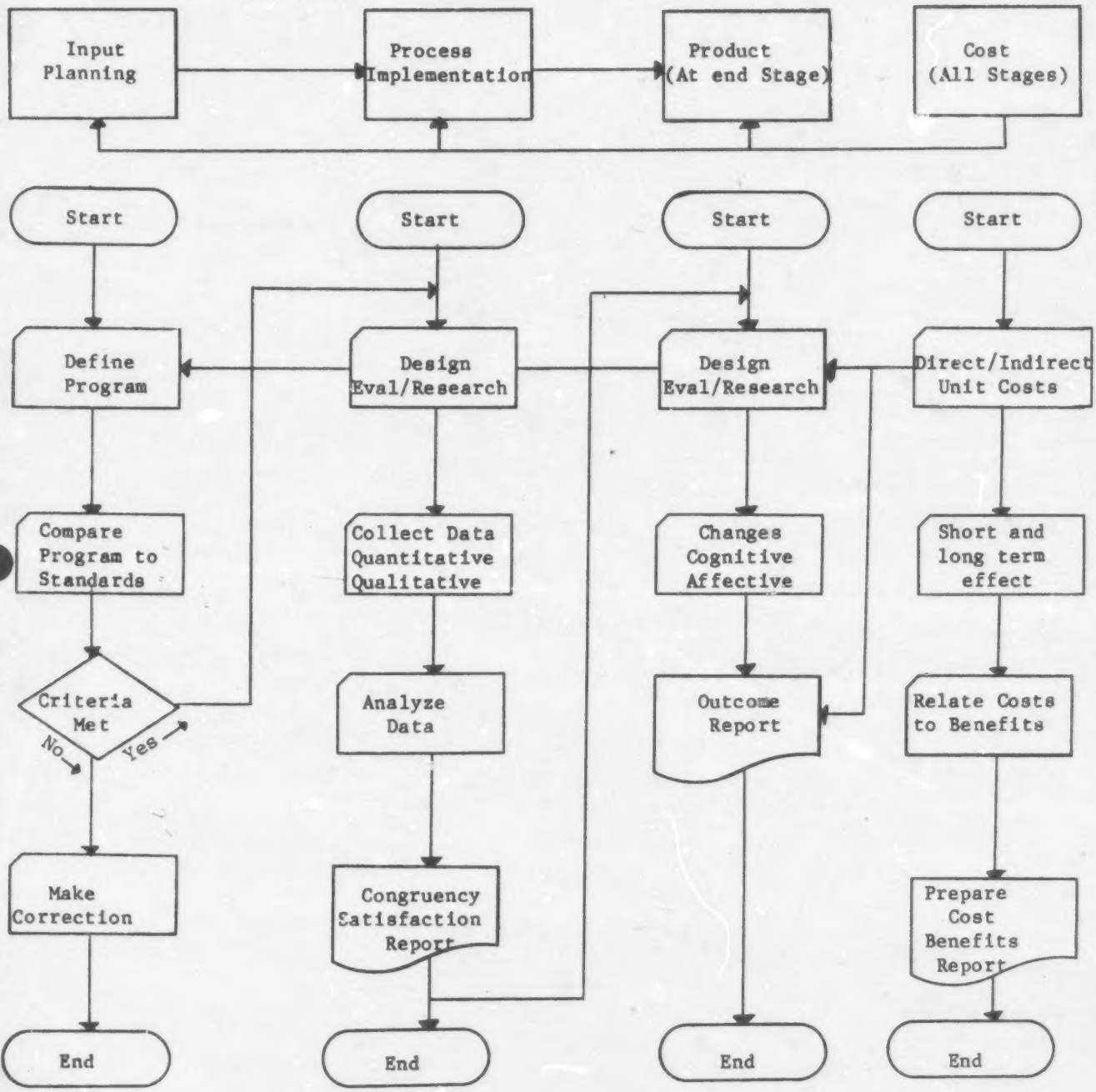
1. The input evaluation of the planning stage determines adequacy of program objectives and goals, resources, and timelines for implementation to meet documented needs.
2. The process evaluation monitors and adjusts the program procedures and activities to facilitate achievement of the previously determined program goals,
3. The product evaluation determines efficiency and effectiveness of program outcome.
4. The cost benefit evaluation at both the planning/input stage, at the process stage, and at the end of the evaluation determines projected and actual program costs and benefits based on perceived needs..

During the planning stages of a program, a cost analysis is a required procedure for the budget allocation. Direct and indirect unit costs for each program area are important indices for evaluating the adequacy of the budget. Cost evaluation (audit) during the process stage insures that funds are being expended according to the agreed plan.

At the end of the program, each program cost can be compared to some selected output or product to determine the total cost or unit cost. Administrators make decisions on the allocation of resources based on this information.

Figure 11

A FLOW CHART FOR THE EVALUATION PROCEDURES



SUMMARY

As stated earlier, the MSDE conception framework being used at the State level follows the same steps as identified by NASDE (1980) in Figure 12 for conducting a specific evaluation study in response to information requested by key decision maker, i.e., Federal and State legislators, State Board of Education members, etc.

In addition to three evaluation questions specific to Bylaw 13.01.04, MSDE addresses the same major quesitons that are addressed annually in the Federal Office of Special Education Annual Evaluation Report to congress (1979) in order to have compatible data. Maryland also uses a discrepancy model (which is consistent with NASDE and OSE) to determine the variation between a LEA Program Status and agreed upon Statewide standards.

In summary, it is important that Special Education Administrators always remember to refer to the National Evaluation Standards to assist them in reviewing any model or Special Education study as a way of judging the structural adequacies of the evaluation component.

This chapter has attempted to outline the ongoing system evaluation cycle that MSDE uses from program planning through the final LEA Project Report or post site visit monitoring and evaluation report. Figure 12 displays the steps that MSDE follows when conducting a specific evaluation study in response to a request by key decision makers at the State level. Therefore, there are ongoing LEA Evaluation and special requested evaluation studies that make up MSDE's statewide evaluation responsibilities.

The next chapter will describe procedures to assist LEA and SEA staff in planning and selecting a product evaluation design for a specific direct service program area.

Figure 12*

STATE OF THE STATE REPORTING
MANAGING BY INFORMATION AND STATEWIDE EVALUATION

INFORMATION FOR EVALUATION AND DECISION-MAKING

STEPS

1 IDENTIFY GOALS, STANDARDS, IDEALS, ESTABLISH QUESITONS

2 DEVELOP SUB-QUESITONS, ETC.

3 ANALYZE EXISTING INFORMATION CAPABILITY AND INFORMATION HOLES

- Monitoring Data
- LEA Report Data
- End of Year Reporting Data
- Fiscal Reports
- LEA Applications
- Discretionary Programs
- Staff Activities
- Other

DETERMINE METHODS OF FILLING INFORMATION HOLES

- Research/Evaluation Studies
- New Forms
- Adapt Existing Forms
- Other

COLLECT INFORMATION

DEVELOP REPORT(S)

- For Legislature
- Staff Performance
- State Board
- Statewide Dissemination
- Other LEA

DECISION-MAKING
MANAGEMENT BY INFORMATION

- Program Development Priorities
- Monitoring Changes
- Policy Changes
- T.A. Training Emphases
- Information Dissemination
- Other

START OVER

* NASDE 1980

CHAPTER V

GUIDELINES FOR DEVELOPING A PRODUCT EVALUATION DESIGN

This chapter's purpose is to describe or select a product evaluation design for a program. It provides a frame of reference to assist the LEA and MSDE staff in selecting which evaluation design is best suited for illustrating that a program works.

LEAs are encouraged to use the format developed by the Joint Dissemination Review Panel (JDRP), established by HEW in 1972, to review products and practices developed in federally supported programs, and decide which ones merit the endorsement of the Education Division. The chosen projects are designated "exemplary", and school systems throughout the nation are urged to look them over and try them. Federal funds are used to support dissemination activities.

To earn the JDRP's seal of approval, you must persuade the panel that your program is effective. The panel is made up of 22 experts in education and evaluation, 11 from the National Institute of Education and 11 from the Office of Education. Their criteria represent one important set of government standards for judging success in educational interventions.

The methods (Phases) used in the JDRP process are summarized below:

Phase 1: Did a change occur?

The first step is to prove that something changed as a result of the program. A change can be claimed for any outcome or consequence that is logically related to the innovation.

To prove that a change occurred, you must have valid and reliable measures of conditions after the intervention has had time to take effect. You must also have estimates of what would have happened at the same point in time if there had been no intervention at all.

Phase 2: Was the effect consistent enough and observed often enough to be statistically significant?

To be convincing, you either have to try educational interventions on large numbers of people or repeat them often enough with small samples to yield meaningful evidence. If gains are small or difficult to observe, you will also have to use larger samples or repeated measurements.

Phase 3: Was the effect educationally significant?

Educational significance, unlike statistical significance, is not only a technical question but a matter of judgment. Educational significance is sometimes established on the basis of past experience, sometimes on expert review, or on theory or statistical rules of thumb. In using past experience, you rely on the results of previous programs. If you use expert review and theory, you might ask specialists to offer their assessment of the project's accomplishments. Statistical rules of thumb usually link educational meaning to statistical significance, sometimes equating the two.

Phase 4: Can the intervention be implemented in another location with a reasonable expectation of comparable impact?

First, you must provide evidence that your projects' setting is not unique. For example, if your high school is near a computer firm, you might easily arrange study programs where students can learn keypunching and software design. The same program would be difficult to repeat, however, in a school without access to a computer company.

Second, you should be able to show that your staff is not unique. If only one teacher has used the program, it is hard to tell how much of a gain can be credited to the program and how much to good teaching. The same is true even if the teacher has repeated the program successfully several times.

Third, the JDRP wants something to show that the intervention can probably be duplicated. Most evaluations do not pay much attention to proving whether a program might be useful to others under different circumstances, even though some methods for establishing generalizability are available and would strengthen the evaluation design.

Phase 5: How likely is it that the observed effects resulted from the intervention?

A number of explanations can be offered for any changes that appear to be associated with a program.

Alternative explanations are particularly important when you do not use rigorous experimental designs. Since the JDRP requires evidence that improved performance is a result of your particular educational intervention, you will have to rule out the possibility that other factors were actually responsible. Some of the factors that can be confused with program performance are history, maturation, statistical regression, mortality, instrumentation, testing, selection, the reactive effects of testing, the Hawthorne Effect, and interference from other programs.

Phase 6: Is the presented evidence believable and interpretable?

The sixth and last hurdle has to do with the general quality of the evaluation evidence you present to the JDRP.

- Present a complete description of the evaluation's methodology and findings.
- Describe instruments in detail. Document reliability and validity information, the number of students tested, testing dates, and the level of forms of the tests used.
- Present data for all of the measures administered. Give types of scores (e.g., norms or raw scores). Clearly identify statistical tests and provide sample sizes each time scores are reported.
- Be sure the inferences you draw from analyses are consistent with the evaluation's design and information.
- Be objective in your presentation, and always describe any limitations on the evaluation's design.

The methods (Phases) utilized in the JDRP process can be applied to many programs and are presented in the JDRP's Ideabook, prepared by Dr. G. Kasten Tallmadge, in the resource component of this kit.

A general procedure for selecting a product evaluation design for a program area is identified in the following six steps.

1. Define goals and objectives of the evaluation (input, process, product, cost);
2. Develop evaluation standards;
3. Define population;
4. Select samples;
5. Select evaluation design;
6. Select or develop data gathering instruments; and
7. Planning the evaluation report.

Step 1. Define goals and objectives of the evaluation

The following questions must be answered about the program to be evaluated before a design can be selected or developed:

1. What are the stated objectives of the program?
   Staff awareness? Changes in behavior? Client satisfaction?

2. What is the target population of the program to be evaluated?
   Students? Parents? Teachers? Organizations?

3. What is the timespan with which the program to be evaluated is concerned? Is the program primarily interested in immediate results, in long-range change, or simply maintenance?

4. What are the means of achieving the goals of the program to be evaluated? Who is to carry out the program? What will they do? How shall they measure their success?

Based on the answers to the kinds of questions above about the program to be evaluated, the goals and objectives of the evaluation design can be developed. These goals and objectives will address the particular specifications of those aspects or variables of the program that have been selected to be evaluated. The goals and objectives must describe those observable behaviors or events that the evaluation design, when selected or developed, claims to measure.

Step 2. Develop evaluation standards

Evaluation standards are developed through the process of deciding the type of information that will provide substantial evidence of a program's success. A program is usually considered successful if it accomplishes its goals and the activities used are beneficial. Standards are required to encompass a particular view or a specific set of criteria with which to judge educational programs and projects. There are several ways to set standards: (1) use specialists, (2) rely on past performance to determine reasonable expectations, (3) rely upon measures of improvement, and (4) rely upon established practices and norms.

When the evaluator has decided what information will be collected and the standards to be used to judge the information, an evaluation design must be selected or developed to compile the information and insure that the information is both valid and reliable. See Appendix __ for a detailed review of Standards for Evaluations of Educational Programs, Projects and Materials, supported by the National Institute of Education and others.

Step 3. Define population

The population or target group is that group about whom the evaluator is interested in gathering information and drawing conclusions. This is termed the population of the evaluation. The first step in sampling, then, is to define this population.

In defining the population to be sampled, the evaluator establishes boundary conditions which specify who shall be included in or excluded from the population being studied. Some studies may require a total population to be included.

Step 4. Select the sample

The most commonly used sampling procedures, briefly explained here, offer the greatest likelihood that the sample population selected represents the population being studied.

Simple Random Sampling A simple random sample is one in which each individual in the population being studied has an equal chance of being included in the sample population. To secure a random sample, tables of random numbers are often used. Such tables are available in the appendices of most statistical texts.

Stratified Random Sampling Stratified random sampling utilizes supplementary information about the population to draw samples that are more apt to represent the population.

The use of stratified random sampling procedures permits the evaluator to include parameters of special validity, in terms of selection factors, through the use of moderate or control variables. In addition, stratification is a useful operational strategy for selecting members of the population for the study and for reducing the variability of the sample.

The first step in stratified sampling is to identify the stratification parameters or variables. Each stratification parameter represents a control variable, that is, a potential source of error or extraneous influence which may provide an alternative explanation for the outcome of a study.

While randomness is the key to overcoming selection bias in sampling, stratification adds precision by insuring that the sample contains the same proportional distribution of respondent, on selected parameters as the population. If a non-random sample were to be drawn, its correspondence to

the target population would not be insured. Stratification increases the likelihood that the sample is representative of the population.

When stratified sampling is used, it is important to insure that, within each stratum, sample respondents are chosen randomly from the population. The procedure enhances the likelihood that selection sources of invalidity other than those controlled through stratification are eliminated. Since stratification controls for selection invalidity based on the systematic application of pre-selected variables, it is recommended for use by evaluators with those variables identified as representing the greatest source of potential selection bias.

Determining Sample Size The primary issue in choosing a sample size is to insure that the sample selected is of sufficient size to represent the population from which it is drawn. While an evaluator can never be certain that the sample selected is truly representative unless the entire population is tested, he can establish that the sample is representative of the population, within critical parameters, at an acceptable level of probability. This probability level is called a confidence level and is usually set at 95% (the so-called .05 level), meaning that there is a 95% chance that the sample is distributed in the same way as the population.

In stratified sampling, an evaluator attempts to maintain the sample proportionality on all stratification parameters in the sample that occurs in the population.

Example:

Population: All teachers in Allegany County (Variables Are Controlled by County Stratification)

Positions:	
Special Education Professional	12%
Special Education Paraprofessional	10%
General Education Professional	60%
Related Services Professional	18%

$$N = (z/e)^2 (p) (1-p)$$

N is the sample size, z is the standard score corresponding to a given confidence level, e is the proportion of sampling error in a given situation, and p is the estimated proportion or incidence of cases in the population.

Allegany County Special Education Professionals account for 12% of all county professionals, or, as in this example, (p = .12). For a confidence level of 95% (z in this example = 1.96) with a tolerable amount of error no greater than plus or minus 10%, the necessary sample size would be calculated as shown below:

$$N = (1.96/0.10)^2 (.12) (.88)$$

$$N = (1.96)^2 (.1056) = 384.16 \times .1056 = 40.57$$

Thus, a sample of 41 special education professionals will create no more than a plus or minus 10% sampling error with a confidence limit of 95% for a population with the incidence of 12% special education, general education and other categories of professionals. (Values for z available in statistical texts.)

When there is some question about sample size, as large a sample as practical should be selected, since with a large sample, the sampling error is likely to decrease. In addition, large samples add to the stability of the findings obtained.

Step 5. Select an evaluation design

The purpose of evaluation is to provide a means of making valid decisions and judgments about a program's effect on the real world. Each product evaluation design has features that guard against certain biases. True-experimental design is the most trustworthy. The evaluator should be confident that his or her design insures maximum internal validity (that is, observed changes must be due to the program and not to outside events). Not all biases affecting validity are equal in a given situation, nor are all evaluation questions of equal importance. A good evaluation design controls most, if not all, of the likely threats to internal validity. However, this control alone does not insure that the program caused the observed changes. The only way to establish causality is with a true-experimental design that compares two or more programs and assigns subjects randomly to treatment and control (placebo) groups.

Experimental design

This type of evaluation design uses experimental and control groups. Out of the target population, units (people, work teams, classrooms, or schools) are randomly chosen to be in either the group that receives The program being evaluated or the "control group" which does not. Measures are taken of the relevant criterion variable before the program starts and after it ends. Differences are computed, and the program is deemed a success if the experimental group has improved more than the control group. See Figure 13.

Figure 13

Randomized Control Group Pretest Post Test Experimental Design

Group	Before	Treatment	After
Experimental	a	x	b
Control	c		d

If the difference between a and b is greater than the difference between c and d, the program is a success.

It is not necessary for the control group to receive no program. Such a condition would be important only if the decision to be made is a choice between the program in question and no program at all. More often, the choice will be between the experimental program and the usual treatment or program for the group, in which case the control group would receive the standard treatment or program against which the experimental treatment or program would be measured. Pre-intervention measures are not a necessary condition in an experimental design. Post-intervention measures will be sufficient if the selection of experimental and control groups has been truly random.

A basic issue in selecting a research design is to fit the design to the purpose of the study. When the purpose of the study is to show causality, then the essential requirement for the true-experimental design, the random assignment of people to experimental and control programs, must be met. Evaluators suggest that this is often more possible in the real world than many suspect. When resources are scarce and some people must do without the program or service, a random selection of the recipients and non-recipients may be possible; when new programs are introduced over a

period of time, the later recipients can become the control group for those who receive the programs early. For these reasons, the evaluation of pilot projects can often be designed on an experimental basis.

Decisions about the continuation or abandonment of the program and decisions on whether to advocate nationwide use of the program model require great confidence in the validity of the evaluation and therefore require an experimental design. Other types of decisions may not need such a rigorous evaluation design.

If the evaluator wants information for day-to-day management purposes, a good automated records system, supplemented by periodic follow-up measurements of outcome, is likely to be the most useful. If the evaluator wants "formative" information on the program (example: process evaluation), only rough estimates may be needed. These estimates can be taken from the real program as it shifts and changes. In this case, simpler techniques (cheaper, faster, easier to manage, and less demanding of program staff than true-experimental designs) can be used.

Weiss (1972) suggests that a preliminary evaluation using soft techniques (correlational, ex-post-facto, one group pretest post-test) can show whether a program warrants further evaluation.

Programs are rarely operated for the convenience of the evaluator. His/her requirements are only occasionally a factor in the program (and funding) arrangements. For these reasons, there is a trend toward quasi-experimental design.

Quasi-experimental design

Quasi-experimental designs, those that do not satisfy the strict requirements of the experimental design, have been given a boost toward legitimacy by the work of Campbell and Stanley. The basic criterion for

judging the relative value of quasi-experimental designs is the extent to which these designs protect against the effects of extraneous variables on the outcome measures. The best designs are those that control relevant outside effects and lead to valid inferences about the effects of the program.

Quasi-experimental designs are in no sense poorly constructed experimental designs. They have a form and logic of their own and have the advantage of being practical when conditions prevent the use of true experimental designs. Recognizing in advance what quasi-experimental designs do and do not control for and the misinterpretations of results that are possible, one may draw a number of conclusions.

The designs described here have been adopted from the widely used description of quasi-experimental designs provided by Campbell and Stanley (1967). In addition to naming and describing the designs, each design will be presented graphically.

In these instances, the term treatment refers to a program, policy, or practice being evaluated. "Measurement" is used in the widest possible sense of terminal behavior in the cognitive or attitudinal domain.

Design 1, the One-Shot Case Study. This design involves the administration of measuring devices to a group of learners having received the same sort of treatment, as shown in the following paradigm. See Figure 14.

Figure 14

The One Shot Case Study Design

	Before	Treatment	After
Group Tested		Yes	Measurement

This design is useful for a simple formative evaluation, perhaps in the early stage of a more thorough evaluation design, to obtain some indication of what happens to learners after they experience a certain kind of treatment.

Design 2, the One Group Pretest Post-Test Design. This design consists of pretesting and then post-testing a single group who, between the two testings, has received the treatment in question. See Figure 15.

Figure 15

The One Group Pretest Post-Test Design

	Before	Treatment	After
Group Tested	Measurement	Yes	Measurement

The technical defects of such a design are numerous, for it is particularly susceptible to such threats to validity as history, maturation, testing, instrumentation, plus the possible interaction of selection with other factors.

The design is useful to formative evaluators, however, who may wish to gather data regarding the effectiveness of particular subcomponents of an instructional program.

Design 3, the Nonequivalent Control Group Design. The least satisfactory of the evaluation designs involving a control group is one in which the assignment to experimental and control groups is not accomplished randomly. The nonequivalent control group design employs two or more groups, each of whom receives a pretest and, following an instructional

treatment, a post-test. In its simplest form, with one group serving as an untreated control group, the design is seen in Figure 16.

Figure 16

The Nonequivalent Control Group Design

	Before	Treatment	After
Group 1	Measurement	Yes	Measurement
Group 2	Measurement	No	Measurement

The major applications of this design are to those situations in which it is impossible to constitute comparison groups randomly. Although the advantages of randomization cannot be obtained, many of the dividends involved in the control of comparison groups are present. The more similar the comparison groups, the more straightforward are the interpretations of the data collected via this design.

The need for this particular design arises often in the practical school world where randomization cannot be employed.

Design 4, the Pretest Post-Test Control Group Design. When randomization is possible, then a stronger design, the pretest post-test control group design, is available. This design controls for almost all eight threats to a study's validity cited by Campbell and Stanley.

In its most basic form, this design can be depicted as seen in Figure 17.

Figure 17

Pretest Post-Test Control Group Design

Group	Before	Treatment	After
Random Experimental	Measurement	Yes	Measurement
Random Control	Measurement	No	Measurement

This design can be varied so that there are two or more treatment conditions. The untreated control group may or may not be employed. This design is particularly useful because of the ease with which the data collected can be interpreted. The only real defect in this design is that if the pretest is reactive (when completed by learners, it influences their response to the treatment condition), its influence may confound the findings.

Design 5, the Post-Test Only Control Group Design. When randomization is possible, yet there is concern that the pretest may be reactive, as described above, a particularly useful design is the post-test only control group design, as seen in Figure 18.

Figure 18

Post-Test Only Control Group Design

Group	Before	Treatment	After
Random Experimental		Yes	Measurement
Random Control		No	Measurement

This design is the same as the pretest post-test control group design, except that there is no pretest.

The basic benefit of the post-test only control group design is that by measuring an untreated, randomly assigned control group the evaluator obtains an estimate of how the treated control group would have responded on a pretest, without introducing the potentially reactive effects of the pretest.

This design can be strengthened by using a control group which is as similar to the treatment group as possible, such as using students in the same schools for each group.

An even better course is to expand data collection procedures to those of a quasi-experimental design. Starting with post-test measures of one group, the evaluator can take similar measures of a new group entering the program. The post-tests of the first group can be compared with pretests of the next group; pretests of this group can then be compared with their post-tests, as seen in Figure 19. In this fashion, more valid conclusions can be drawn.

Figure 19

Post-Test Control Group with Pre-test and Post-test Design

Group	Before	Treatment	After
Random Control		No	measurement
Random Control	measurement	No	measurement
Random Experimental	measurement	Yes	measurement

Design 6, the Interrupted Time-Series Design. The interrupted time-series design can be diagrammed as follows:

Figure 20

Time-Series Design

Before	Treatment	After
M_1 M_2 M_3	Yes	M_4 M_5 M_6

The design is a series of measurements (the more, the better) taken both before and after the introduction of the treatment. Frequently these measurements are drawn from existing records, such as regularly administered achievement tests, attendance reports, or disciplinary referrals. Such so-called archival data are particularly well suited to the interrupted time-series design but are by no means the only kinds of data that can be used with the design.

The interrupted time-series design is one of the most attractive quasi-experimental designs. It involves a series of measurements taken at periodic intervals beginning before the program intervention is initiated,

continuing through the program, and concluding after the program intervention ends. Such a measurement sequence allows an evaluator to determine whether the measures immediately before and after the program intervention are a continuation of earlier patterns or whether they mark a decisive change. The interrupted time-series probably protects against all threats to validity except history. The greatest advantage of time-series designs is not that they offer an alternative to traditional comparative designs, but that they provide a markedly different perspective for evaluating the effect of a program intervention over time.

There is a temptation, when describing evaluation designs, to cite many more than the six basic designs described above. There are, of course, many other designs that an evaluator might want to call upon in certain situations. The designs cited here should, however, meet an evaluator's requirement in the majority of situations.

For those readers wishing to become better versed in evaluation design, the reference works by Airasian (1974), Berber (1974), Campbell (1963), Poplam (1975, Olsen (1979), Weiss (1972), and Plogen (1971) are cited at the end of this chapter.

Step 6. Data gathering instruments

One of the major problems of any evaluation program is the selection of data gathering instruments. There are many issues that must be considered in selecting data gathering instruments, and the selection should enhance the individual goals of the program and the procedures that are to be used.

The instrument selected must take into account all possible independent variables and dependent variables. An independent variable is referred to as an input or stimulus factor which effects the output. A

dependent variable is a response by the recipient to the input or stimulus factor, i.e., an observed aspect of the behavior of the target population is in response to the stimulus factor.

In choosing a data collection instrument to answer a particular question, four factors should be considered. First, the instrument should be agreeable to both the evaluator and the participants in the evaluation. If the evaluator wishes to use questionnaires, but the staff prefers interviews, the consequences of imposing the evaluator's will on the participants must be considered. Second, the data collection instruments should be technically sound and insure that the data collected using the instruments is reliable, valid, and targeted to the evaluation questions. Third, the data collection instruments should provide the best data the evaluation budget can afford. Questions such as whether to buy or develop particular instruments and whether to use more than one type of instrument for each evaluation question will have to be addressed. Fourth, the data collection instruments chosen must allow enough time for gathering and analyzing the data collected.

It is possible to collect evaluation information in a variety of ways. In questioning whether or not an innovative special education reading program has actually improved student reading ability, for example, an evaluator could ask the question: How much have students' reading skills improved?

In answering this question, any of the following information collection techniques could be appropriate:

> Giving parents a rating scale to assess their children's reading performance.
>
> Sending questionnaires to teachers to poll their opinion about student reading performance.

Interviewing students to assess their opinions about student reading performance.

Observing students as they read and rating their reading ability.

Having students keep a diary of their progress.

Giving students a nationally normed achievement test that assesses reading performance.

Having students take a teacher-developed test.

Reviewing student records for criterion-referenced test data (CRT), achievement test scores, report card grades, and teachers comments.

The list above illustrates six alternative types of evaluation instruments an evaluator might use to answer the question posed earlier about improved reading skills. The alternatives illustrated are questionnaires, standard observations, interviews, rating scales, tests, and documenting analysis.

Questionnaires

Questionnaires are self-administered survey forms that consist of pre-determined questions. Answers to questionnaire items can require free responses (short answers) or they can be structured into "forced" choices (multiple choice items). Questionnaires are a relatively inexpensive way to reach large numbers of people quickly. Even when printing and mailing charges are considered, questionnaires cost considerably less than interviews to reach the same number of people.

For the above reasons, questionnaires are frequently used in large scale evaluation designs. The information available from them may be limited, however, since there is always the concern of whether or not the questions have been understood or truthfully answered. As in all data gathering instruments, the reliability and validity of the data gathered through the questionnaire process is of paramount importance. The

construction and design of questionnaires is therefore no easier than that of other data collection instruments.

A major problem with questionnaires is the possibility of a low and perhaps unrepresentative response rate. Recipients may be unable to read or understand the questions, or they may tire before completing all the questions. Many evaluators do not use questionnaires for these reasons, but there is virtually no data gathering instrument without limitations. In any case, the construction and design of questionnaires must emphasize the clarity and reliability of the individual questions in a format that facilitates the completion of the items and the prompt return of the questionnaire when completed.

The mailing of a questionnaire to a sample of respondents typically includes a cover letter, the questionnaire itself, and a stamped return-addressed envelope.

The cover letter is a critical part of the initial mailing because it must establish the legitimacy of the evaluation and/or the evaluator. The cover letter should be brief but should focus on the following points:

The purpose of the evaluation
The protection afforded the respondent
Endorsement of the evaluation
Legitimacy of the evaluator
Opportunity for debriefing
Request for cooperation
Special instructions

After a period of about two weeks to a month has elapsed, it is often helpful to correspond with those respondents who have not yet returned their questionnaires.

Observations

Standard observations require careful planning so that the information obtained is accurate. Observations can give evaluators first-hand information about a program, and are often the only feasible and economical way to gather certain kinds of information. Rater and inter-rater reliability are major considerations in any evaluation design requiring observations, however, and the training of observers (raters) may therefore be a major component in the evaluation budget. Another drawback to the observation process is that people often act differently when they are being observed.

Time sampling observations involve repeated observations of a given situation. Observers may note how many students and which students ask direct and indirect questions during ten consecutive five-minute intervals.

Time sampling allows first-hand observations of a program. The numerous observations required make it possible for the observer to identify unusual events that might otherwise seem to be routine occurrences. When all observations are made in close or virtual sequence, however, they are likely to depict only one particular situation and not represent the program as a whole.

Interviews

An interview is a conversation in which one person tries to obtain information from another. An evaluator may wish to use interviews to learn more about the people who participated in a program and what they thought of the program.

Interviews are both expensive and time consuming but are often invaluable, particularly when information is needed from those who may have

difficulty reading or writing or when investigating complex and sensitive topics or feelings.

The interview method requires trained interviewers. The level of skill required of the interviewer will depend on the nature of the information being collected. The training should include both familiarizing the interviewer with procedures for recording responses and keeping accurate records of all interviews.

Rating and ranking techniques

Evaluators are frequently asked to find out how participants feel about a program. Are students satisfied? Has the staff's morale improved? Which of the three programs did people like best? Did the participants think the program was worth the money?

Although opinion differs concerning the degree of emphasis to place on evaluating attitudes, most investigators of educational issues support the importance of learning what students and teachers who take part in a program think about the program. Evaluation questions that ask about feelings, interests, beliefs, and values are called "attitudinal." Current knowledge in the field of human development suggests that there is an intricate relationship between how people feel about things, what they know, and how they behave. Knowing the facts about good nutrition may not alone convince young people to eat balanced meals, but if they believe that their appearance is important in making friends, they may be more likely to follow wholesome dietary practices.

Rating scales can be used for self assessment or for appraisals of other people, groups, events, or products. Student attentiveness, for example, can be rated on a five-point scale from 1, for not very attentive,

to 5, for very attentive. Rating scales are particularly useful when the evaluator wishes to reduce judgmental data to a manageable form. Rating scales are relatively easy to complete and produce objectified data. Unfortunately, they are subject to many types of bias; some raters are lenient and others are not, and sometimes raters let their personal feelings influence their ratings (a halo effect). Further, the amount of information that can be obtained from rating scales is generally limited because the rating categories are never perfectly descriptive.

The Likert Scale, a type of rating scale, relies on response categories such as strongly agree, agree, undecided, disagree, and strongly disagree. For scoring purposes, strongly disagree is usually assigned a weight of 1, with favorable attitudes reflected in higher numerical scores.

The Semantic Differential. In this procedure the evaluator asks the respondent to indicate his feelings about a particular topic by checking some point on a scale between opposites, such as good and bad. When properly used, the semantic differential will provide information that sorts out negative attitudes from positive ones. People who admire teaching will generally view it as fast, good, valuable, strong, and pleasant, while those who don't will generally choose the opposite adjectives.

The creators of the semantic differential technique (Osgood, Suci, and Tannenbaum) have grouped adjective pairs into three major categories: evaluation, which includes "good-bad" and "fair-unfair"; potency, which includes "strong-weak" and "heavy-light"; and activity, which includes "fast-slow" and "active-passive." Hundreds of studies have supported the

semantic differential to discover people's predominant feelings about a variety of ideas and subjects.

Ranking scales involve ordering a set of items in a hierarchy according to some value or preference, such as asking a teacher to rank four textbooks from one to four according to their reading difficulty. Like rating scales, ranking scales are generally easy to complete and produce objectified data. It should be remembered, however, that rank ordering a long list of items can seem quite tedious and time consuming to the respondent. Ranking scales sometimes ask respondents to make distinctions between or among items that have no apparent differences to the respondent.

Rather than asking people how much they like or dislike something, ranking scales present people with an actual choice between different items. The number of choices usually ranges from two to four.

An evaluator may use such items to determine which students would be happier in a relatively "open" program and which ones favor a more traditional approach. To be valid, the items would have to have been used enough to demonstrate that certain groups of students make their choices in some systematic way.

Forced choice items are relatively difficult to develop. Forced choice techniques may be more objective than the Likert scale, however, because people completing a Likert scale often sense whether agreement or disagreement is preferred by the evaluator. Forced choices are generally simple to score and interpret.

Tests

Performance tests require people to complete a task, after which the quality of the performance or product is assessed. One example of a performance test is a typing test in which a person types from a sample for a fixed period of time. The evaluator is then able to count the number of words typed correctly during the allotted time. The major advantage of performance testing is that it relies on tasks that are close to "real world" activities. It is often very time consuming and expensive, however, because performance tests generally have to be administered individually and often require the use of special equipment.

Criterion-referenced tests (CRT) measure how well an instructional task or skill has been achieved. Criterion referenced tests have:

- Clearly defined educational tasks or objectives.
- Test items that are specifically designed to measure performance on these tasks or objectives.
- Scores that are interpreted in terms of attaining a set criterion or level of competence.

Most CRTs report scores either as the number of items correctly answered for each objective or sometimes as mastery or non-mastery scores, "mastery" meaning correctly answering an arbitrarily selected number of items per objective. These types of score interpretations are sometimes accepted by theorists as legitimate ways of expressing CRT scores and have meaning for teachers who know their curriculum.

The procedures used to validate CRTs are not very sophisticated, and field test results are not reported in any detail. When it comes to the sample's size and representativeness and the amount and precision of the data presented in technical reports, most CRTs cannott compare with the highly structured field tests conducted to validate norm-referenced tests.

From a theoretical point of view, however, there is no reason why CRTs cannot be used in evaluating educational programs, as long as proper development and validation procedures are followed. See Ysseldyke (1978), Assessment in Special Education, for an extensive review of using standardized tests for evaluating handicapped students.

Documentary analysis

Documents and records can provide evaluation with pertinent data for use in answering certain evaluation questions. Documentary analysis, sometimes referred to as "content," "activity," or "informational analysis," is closely akin to historical research. Both methods of investigation require the examination of existing records, but while historical research is primarily concerned with the more distant past, descriptive research is concerned with the present and has much utility for the educational evaluator.

A wide variety of documentary surveys are made. Some evaluators analyze judicial decisions, state laws, and school board rulings. They count and classify specific items in these sources that relate to their problems, such as information about salary schedules, certification, liability for school accidents, and district organization. Evaluators may also gather data describing existing school practices, processes, and conditions from administrative records, forms, and reports, such as committee reports and minutes of meetings, budgets and financial records, and cumulative attendance and health records.

Classroom teachers are generally concerned with identifying the most important skills and knowledge that children should attain. In addressing this concern, many evaluators have turned to documentary research. In the field of mathematics, they have conducted studies to discover what

arithmetic processes are most commonly employed in business and social usage. Wilson and Dalrymple (1937) examined 102,000 uses of fractions which they obtained from business records and concluded that 90 per cent of ordinary adult usage of fractions was confined to halves, thirds, and fourths.

Documentary analysis may help educators (1) describe specific conditions and practices that exist in schools and society, (2) identify trends, (3) detect weaknesses, (4) trace the development of a student's or writer's work, (5) discover differences in the practices that prevail in various areas, states, or countries, (6) evaluate the relationships of stated objectives and what is being taught, and (7) detect the biases, attitudes, interests, values, and psychological states of people. There are many more types of data gathering instruments. See Appendix D - the glossary for additional instruments.

The Evaluation Report

This book has been mainly concerned with designing and initiating a program evaluation. It's also necessary to deal with the results of this process: the evaluation report.

The evaluation report should be easy to understand and should communicate what was done, how it was accomplished, and why. The evaluation report must be credible and must logically describe the Evaluation process, including:

- an introduction to the evaluation, the evaluation questions, and limitations on the scope of the evaluation;
- the design strategy and the sampling procedures for each evaluation question, and their limitations;
- the information collection techniques and instruments, their limitations, and any field activities;

- the methods used to analyze the evaluation information, their limitations, and the results for each analysis;
- the answers to each evaluation question, including interpretation of findings and recommendations.

The introduction should briefly describe the program being evaluated, the group that is conducting the evaluation, and the approach to evaluation. Any legislation that created the program and mandated its evaluation should be covered. The process used to arrive at statements for program goals, activities, and evidence of program merit should be included. It should also explain the choice of evaluation questions, list the questions, and describe any constraints imposed upon the scope of the evaluation.

The evaluation's design and sampling procedures with any limitations should be described. There should be a justification for how participants were grouped, variables that would affect results, and any limitations on external and internal validity. The sampling procedure should be noted along with any subdivisions into which participants were divided. Problems encountered in selecting participants should be mentioned.

The report should include a description of instruments used and the persons to whom administered with an opinion as to test reliability. It is also a good idea to explain how instruments were administered and scored. Field activities should be summarized with notes of any irregularities that might affect the findings.

A description of the relationships between separate analytic methods to evaluate questions should be included, with method and design sources noted as necessary. The answers to evaluation questions should be clearly tabulated and labeled. Any recommendations included with data should focus on program improvement. Many evaluators provide only generalizations that

are simply too sweeping to be practical. If recommendations are vague or open to various interpretations, clients will not be able to translate the evaluation into effective action.

SUMMARY

Professionals wishing to use rigorous evaluation methods to answer educational questions are immediately confronted with numerous detailed problems which have to be worked out for each new project. There are no general formulas and only a few words of advice and warning. The purpose of the project must be clearly stated; care should be taken to define the population to be examined and the units or characteristics which are to be measured. Once these matters are clarified, the evaluator is confronted with data collection problems. The educational evaluator is usually fated to do a large part of his work with data collected by others under a variety of conditions which cause these data to be of highly variable quality. The evaluator must, therefore, watch for and appraise errors of estimation, fictitious statistics, series built up through the use of substitutes, classification errors, discontinuities, self-interest errors, and definitional errors.

If an evaluator collects his own data directly from the population, he will very often be confronted with faulty memories or resistance to his inquiries. He must describe the information wanted and record it through the imperfect medium of non-technical language where vague or poorly defined words, regional differences in word usage, loaded words, the type of question (open, two-way, or multiple choice), and the order in which the questions are presented, may all influence the information received. Where possible, it will always be better to observe and record behavior directly rather than to ask people to talk about their behavior.

The way in which these difficulties are met and solved in each individual evaluation project will determine whether the findings of the project will make the hoped-for contribution toward the solution of the problems the administrator had sought to solve.

LEAs which are beginning to design evaluations for specific studies should use this resource book as a guide. Assistance can be provided from the Information Management and Program Administration and Evaluation Branches, Division of Special Education, Maryland State Department of Education.

APPENDIX A

Resource Manual References

REFERENCES

Abend, Allen C., Bednar, Michael J., Froehlinger, Vira J., and Steinzler, Yale; Facilities for Special Education Services, A Guide for Planning New and Renovated Schools. Maryland State Department of Education, 1979

Adelman, Howard S. " Evaluating Training Programs in Special Education." 1970. 14 pp. (ED 073 094).

Airasian, Peter W., "Designing Summative Evaluation Studies at the Local Level." In Evaluation in Education: Current Application. Berkeley, California: McCutchem Publishing. 1974

Alkin, Marvin C. "Evaluation Theory Development." Evaluation Comment. 1969, Vol. 2, pg. 2-7.

Alkin, Marvin C. "Wider Context Goals and Goal-based Evaluation." Evaluation Comment. 1972, Vol. 3:4, pg. 5-6.

American Coalition of Citizens with Disabilities, Inc., Final Report of Supplement to Consumer Involvement in Career Education and Vocational Education. 1979.

Anderson, Richard D. and R.J. Sonnen. "Evaluation: Relating Special Education Programs to School Objectives," Journal of Special Education, 7 (Summer 1973), pp. 197-203.

Anderson, Scarvia B., Samuel Ball, Richard T. Murphy and Associates. Encyclopedia of Educational Evaluation. San Francisco: Jossey-Bass, Inc., 1975.

Ashburn, A. G. "Directing education research training toward needs of large school districts." Texas A & M University, Office of Educational Administration.

Baglin, Carol Ann. "Survey of Status of LEA Evaluation of Special Education Within Maryland." Unpublished Report, Maryland State Department of Education, 1979.

Berber, Theodore Xenophon. "Pitfalls in Research: Nine Investigators and Experiments Effects." Second Handbook of Research on Teaching, edited by Robert M. W. Travers. Chicago: Rand McNally. 1973

Blackhust, Eric W.: "Program Evaluation Through Learner Change: Considerations and Applications." Colorado Journal of Educational Research. 16,4, 1977. pp 1-5.

Bloom, B.S.; Englehart, M.D.; Furst, E.J.; Hill, W.H.; and Krathwohl, D.R. Taxonomy of educational objectives: Handbook I: Cognitive domain. New York: David McKay, 1956.

Borg, Walter R. and Meredith D. Gall. Educational Research: An Introduction. New York: David McKay, 1963.

Broger, Barton P. "Program Evaluation" Journal of Learning Disabilities. (June-July 1971).

Bureau for Education of the Handicapped. Summary of RRC Director's Meeting. Lexington, Kentucky: Mid-South Regional Resource Center, 1979.

Campbell, D.T. and Stanley, J.C. "Experimental and quasi-experimental designs for research of teaching." In N.L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Center for the Study of Evaluation UCLA Graduate School of Education. CSE Elementary School Test Evaluations. Los Angeles, California, 1970.

Clark, Kenneth. Dark Ghetto. New York: Harper and Row, 1965.

Coleman, J.S. Equality of educational opportunity. Washington, DC: U.S. Department of Health, Education, and Welfare; Office of Education, 1966.

Cook, D.L. Program evaluation and review techniques, applications in education. U.S. Office of Education Cooperative Monograph, 1966, 17, (OE-12024).

Cronbach, L.J. "Course improvement through evaluation." Teachers College Record, 1963, 64, 672-683.

Cronback, L.J., and Snow, R.E. Individual differences in learning ability as a function of instructional variables. Stanford, CA: Stanford University Press, 1969.

Department of Health, Education and Welfare, Office of Education, Bureau of Education for the Handicapped. Developing Criteria for the Evaluation of Due Process Procedural Safeguard Provisions. 1978.

Department of Health, Education and Welfare, Office of Education, Bureau of Education for the Handicapped. Developing Criteria for the Evaluation of Individual Education Program Provisions. 1978.

Department of Health, Education and Welfare, Office of Education, Bureau of Education for the Handicapped. Developing Criteria for the Evaluation of the Least Restrictive Environment Provision. 1978.

Department of Health, Education and Welfare, Office of Education, Bureau of Education for the Handicapped. Developing Criteria for the Evaluation of Protection in Evaluation Procedures Provisions. 1978.

Dunst, Carl J. "Program Evaluation and the Education for All Handicapped Children Act." Exceptional Children, 1979, Vol. 46:1, pg. 24-31.

Ebel, R. L. Measuring Educational Achievement. Englewood Cliffs, NJ: Prentice-Hall, 1965.

Educational Research: Limits and Opportunities. Third Annual Report of the National Council on Educational Research, The National Institute of Education, U.S. Department of Health, Education, and Welfare, March, 1977.

Eisner, E.W. "The perceptive eye: Toward the reformation of educational evaluation." Invited address, Division B, Curriculum and Objectives, American Educational Research Association, Washington, D.C., March 31, 1975.

Evaluation Technologies Program Center for the Study of Evaluation, Humanizing Learning Program Research for Better Schools, Inc., CSE-RBS Test Evaluations: Tests of Higher-Order Cognitive, Affective, and Interpersonal Skills. 1972.

Evaluative criteria 1960 Edition. National Study of Secondary School Evaluation, Washington, D.C., 1960.

Fink, Arlene and Kosecoff, Jacqueline, An Evaluation Primer, Washington, D. C., Capitol Publications, Inc., 1978.

Flynn, Donald L., RMC Research Corporation, prepared for Educational Testing Service, Princeton, N.J., Report UR-231, An Evaluation of the Cost Effectiveness of Alternative Compensatory Reading Programs, Final Report in Four Volumes, Volume II: Model Sensitivity. 1976.

Gallagher, James J. and Julia Hall. The Benefits of Competency Testing for the Exceptional Student. Education Unlimited. Educational Resources Center, Boothwyn, Pennsylvania, Vol. 1, No. 1, June, 1979, pg. 71-72.

Galloway, James R., et. al. Developing the Compliance Monitoring System in Special Education: A Process Manual for State and Local Education Agencies. Washington, D.C.: National Association of State Directors of Special Education, 1977.

Gephart, William. Evaluation: Past, Present, and Future. Occasional Paper 17. Bloomington, Indiana: Center on Evaluation, Development, and Research, no date.

Glass, G. V. Design of evaluation studies. Paper presented at the Council for Exceptional Children Special Conference on Early Childhood Education, New Orleans, December, 1969.

Glass, G. V. and Maquire, T. O. Analysis of time-series quasi-experiments. U.S. Office of Education Report No. 6-8329. Boulder, CO: Laboratory of Educational Research, University of Colorado, 1968.

Glass, Gene V.; Victor L. Willson; and JOhn M. Gottman. Design and Analysis of Time-Series Experiments. Boulder, Colorado: Laboratory of Educational Research, University of Colorado, 1972.

Grotsky, Jeffery N. and others. Guidelines "PEP," Peer Evaluation Program. "A Systematic Approach for Evaluating Educational Programs, 1973-1974." Harrisburg, Pennsylvania: State Department of Education, 1973. 40 pp. (ED 083 304).

Guba, Egon G., Jeri M. Ridings, and Daniel E. Stufflbeam. "A Progress Report on the Project to Develop Standards for Educational Evaluation." A paper presented at the annual conference of the AERA and NCME. San Francisco, April, 1979.

Guba, E. G. "A study of Title III activities: Report on Evaluation" National Institute for the Study of Educational Change, Indiana University, October, 1966. (mimeo)

Guba, E. G. Alternative perspectives on educational evaluation. Keynote speech at the Annual Meeting of the Evaluation Network, St. Louis, MO, 1976.

Guba, E. G. "Toward a methodology of naturalistic inquiry in educational evaluation." CSE Monograph Series in Evaluation, Los Angeles: Center for the Study of Evaluation, 1978.

Hammond, R. L. Evaluation at the local level. Tucson, AZ: EPIC Evaluation Center, 1974.

Hastings, T. A portrayal of the changing evaluation scene. Keynote speech at the Annual Meeting of the Evaluation Network, St. Louis, MO, 1976.

Hayden, David L.; Clemson, Barry; Streissguth, William O.; and Cook, Blair: Project Monitoring: A Systems Manual, (1974). 14 pp. included as Appendix D.2. in Penn. RRC proposal submitted in 1974.

Hayden, David L.; Olsen, Kenneth R.; and Streissguth, William O.; Program Cost Analysis System Manual. 28 pp. Included as Appendix D.1. in Penn. Regional Resource Center Quarterly Report to BEH for FY 1974-1975.

Hayden, David L.; Steissguth, William O.; Clemson, Barry; and Cook, Blair: Time Accounting System: A Systems Manual. (1974). 9 pp. including sample forms. Included as Appendix D.3. in Pennsylvania RRC progress report to BEH submitted 1974.

Holowenzak, Stephen P. "An Evaluation of the Project STAR Reading Program Intervention." Vols. I. II, and III. Baltimore, Maryland: Maryland State Department of Education, 1979.

Holowenzak, Stephen P., Hayden, David H., Harper, James. "A Sampling Design Matrix for the Evaluation of Individual Assessments and IEP Documentation." Baltimore, Maryland: Maryland State Department of Education, 1979.

Holowenzak, Stephen P. and Stagmer, R.A. "The Cost of Being Accountable: An Objective Referenced Program Cost Model for Educational Management - A Maryland Perspective." Denver, Colorado: The Cooperative Accountability Project, 1976.

Jakowski, E. A study of educational evaluation in the seven largest urban school districts in Texas. Unpublished doctoral dissertation, Texas A & M University, 1976.

Holloway, William H., Eros, Dawn S., Kerr, Michael E., Policy Analysis: An Assessment of Public School Educators' Concerns in Relation to P.L. 94-142. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA., April, 1980.

Jencks, C.; Smith, M.; Adand, H.; Bane, M.J.; Cohen, D.; Gintis, H.; Heynes, B.; and Michelson, S. Inequality: A reassessment of the effect of family and schooling in America. New York: Basic Books, 1972.

Jones, Regionald L., Gottlieb, Jay, "Evaluating Mainstreaming Programs. Models, Caveals, Considerations, and Guidelines." Exceptional Children, May 1978. pp. 588-600.

Kaplan, A. The conduct of inquiry. San Francisco, Chandler, 1964.

Kaufman, R.A. "Toward educational system planning: Alice in Educationland." Audiovisual Instructor, 14, May, 1969, 47-48.

Kershaw, D.N. The New Jersey negative income tax experiment. A summary of the design, operations and results of the first large-scale social science experiment. Dartmouth/OECD Seminar on "Social Research and Public Policies," September 13-15, 1974.

Lessinger, L.M. Every kid a winner: Accountability in education. New York: Simon and Schuster, 1970.

Levin, Henry M. Cost-Effectiveness Analysis in Evaluation Research. In Marcis Buttentag & Elmer Streuring (Eds.), Handbook of Evaluation Research. Beverly Hills, California: Sage Publications, Inc., 1975.

Lilly, M. Stephen, Evaluating Individualized Education Programs: A Premier on Individualized Education Programs for the Handicapped.

Lindquist, E.F. Design and analysis of experiments in psychology and education. Boston: Houghton-Mifflin, 1953.

Lindquist, E.F. (Ed.) Educational Measurement. Washington DC: American Council on Education. 1951.

Linde, James L. and Olsen, Kenneth R. "Minimum Competency Testing and Handicapped Students." Resource Report No. 1, Mid-East and Mid-South Regional Resource Center, January, 1980.

Lord, F.M. and Novick, M.R. Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968.

MacMillan, Donald L., and Semmel, Melvyn I. "Evaluation of Mainstreaming Program" Focus on Exceptional Children. September 1977.

Management Analysis Center, Inc. (Developer and National Association of State Directors of Special Education (Distributer), Special Education Management-By-Information: A Resource Guide for State and Local Education Agencies. (1979).

Martin, Edwin. Progress Toward a Free Appropriate Public Education: A Report to Congress on the Implementation of Public Law 94-142: The Education for all Handicapped Children Act. U.S. Department of Health, Education and Welfare, Office of Education, January, 1979.

McDonald, B. Evaluation and the control of education. In D. Tawney (Ed.) Evaluation: The state of the art. London: Schools Council, 1975.

Meierhenry, Wesley C. Planning for the Evaluation of Special Educational Programs: A Resource Guide. Lincoln, Nebraska: University of Nebraska, 1969. 52 pp. (ED 040 555).

Merriman, H.O. "Evaluation of planned educational change at the local education agency level." Columbus, OH: The Ohio State University Evaluation Center, 1968.

Metfessel, N.S. and Michael, W.B. A paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs. Educational and Psychological Measurement, 1967, 27, 931-943.

Meyer, Judy. The Development of Evaluative Criteria for Training the Handicapped: A Summary of Project Activities. Houston, Texas: University of Houston, Center for Human Resources, 1972. 35 pp. (ED 074 273).

Mowitz, Robert J., Institute of Public Administration, the Pennsylvania State University, The Design and Implementation of Pennsylvania's Planning, Programming, Budgeting System. (NO DATE).

Office of Special Education. Annual Report to Congress on Implementation of 94-142. 1979

Olsen, Kenneth R. Evaluating Effectiveness of Special Education Programs, Mid-South RRC Bulletin #10. 1979.

Ott, J.M. "A decision process and classification system for use in planning educational change." Columbus, OH: The Ohio State University Evaluation Center, 1967. (mimeo)

Parsley, Marilyn J. The Development of a Planning and Evaluating Model for Programs Directed Toward the Exceptional Child in the School System. Doctoral dissertation. Denver, Colorado. University of Denver, 1971. 196 pp. (UM: order no. 71-20,100)

Popham, W.J. Objectives and instruction. In instructional objectives. AERA Monograph Series on Curriculum Evaluation. Chicago: Rand McNally, 1969, 3.

Popham, W. James. Educational Evaluation, Prentice-Hall, Inc. Englewood, New Jersey, 1975.

Provus, Malcolm N. Discrepancy evaluation. Berkeley, CA: McCutcheon, 1971.

Provus, Malcolm, "Evaluation of Ongoing Programs in the New Public School System." In Ralph Tyler (Ed.), Educational Evaluation: New Roles, New Means. Chicago: University of Chicago Press, 1969.

Reinhard, D.L. "Methodology development for input evaluations using advocate and design teams." Ph.D. dissertation, The Ohio State University, 1972.

Report By The Education Advocates Coalition on Federal Compliance Activities to Implement The Education For All Handicapped Children Act (PL 94-142). 1980.

Rice, Joseph M. "The Futility of the Spelling Grind," Forum, XXIII (April, June, 1897). 163-172, 401-419

Rippey, R.M. (Ed.) Studies in transactional evaluation. Berkeley, CA: McCutcheon, 1973.

Sanders, J.R. and Hershiser, M.A. A proposal to study the attributes of a classroom that determine its pervasive quality. Kalamazoo, MI: The Evaluation Center, Western Michigan University, September, 1976.

Scriven, M.S. Evaluation perspectives and procedures. In Evaluation in Education: Current Applications, W. James Popham (Ed.) Berkeley, CA: McCutcheon Publishing Corporation, 1974.

Scriven, M.S. The methodology of evaluation. In R.E. Stake (Ed.), Curriculum evaluation. AERA Monograph Series on Evaluation. Chicago: Rand McNally, 1967, 1.

Scriven, Michael, "Pros and Cons About Goal-Free Evaluation." Evaluation Comment. 1972, Vol. 4, pg. 1-8

Scriven, Michael "The Methodology of Evaluation." Perspectives of Curriculum Evaluation: AERA Monograph Series on Curriculum Evaluation. 1967. Vol 1, pg. 39-83.

Schipp, William and Linda Morra A Conceptual Framework for Developing Statewide Evaluation and MOI Systems in Special Education. Paper presented at NASDSE, Second Annual National Academy. July 14, 1980

Special Education Management by Information: A Resource Guide for State and Local Education Agencies. Management Analysis Center, Inc., Washington, D.C. 1980.

Stake, R.E. Measuring what learners learn. Urbana, IL: Center for Instructional Research and Curriculum Evaluation, 1971. (mimeo)

Stake, R.E. "Nine approaches to educational evaluation." Unpublished chart. Urbana, IL: Center for Instructional Research and Curriculum Evaluation, 1974.

Stake, R.E. "Objectives, priorities, and other judgment data." Review of Educational Research, 1970, 40, 181-212.

Stake, Robert E. "Should Educational Evaluation Be More Objective or More Subjective: More Subjective!" An invited debate at American Research Association Annual Conference. Toronto, Canada, March 1978.

Stake, Robert E. "The Countenance of Education Evaluation." Teacher's College Record. 1967, Vol 68, pg. 523-540

Stufflebeam, Daniel L. "A depth study of the evaluation requirement." Theory into Practice, 5, June, 1966, 121-134.

Stufflebean "The use of and abuse of evaluation in Title III," Theory into Practice, 6, June, 1967, 126-133.

Stufflebeam, Daniel L. Alternative Approaches to Educational Evaluation: A Self-Study Guide for Educators. In W.J. Popham (Ed.), Evaluation in Education: Current Applications. Berkeley, California: McCuthan Publishing Corporation, 1974.

Stufflebeam, Daniel L., et. al. Educational Evaluation and Decision Making. Itasca, Illinois: F.E. Peacock Publishers, Inc., 1971.

Suchman, E. A. Evaluative research. New York: Russell Sage Foundation, 1967.

Taylor, J. P. "An administrator's perspective of evaluation." Kalamazoo, MI: Western Michigan University, Evaluation Center, Occasional Paper #2, October, 1974.

The Policy Research Center, The Council for Exceptional Children, A Policy Research Report prepared for the State of Arkansas, Diplomas, Graduation Requirements and Grading Procedures for Handicapped Students. (1980)

Thorndike, R. L. Educational measurement. Washington, DC: American Council on Educaiton, 1971.

Turner, Ruth M., Macy, Daniel, J., A Five Year Longitudinal Study of the IEP Implementation. Paper. February, 1978, 93p. Ed. 170 983.

Tyler, R.W. Basic Principles of Curriculum and Instruction. Chicago: University of Chicago, press, 1950.

Tyler, Ralph W. "Educational Evaluation: New Roles, New Means." The sixty-eight yearbook of the National Society for the study of Education, The University of Chicago, 1969.

Tyler, R.W. "General Statement on Evaluation." Journal of Educational Research, 35, (1942): 492-501.

Vaughan, R.W., Ed.D. "Management System for Determining Cost-Effectiveness/Cost-Efficiency of Public Schools Special Education Programs." Paper presented at CEC Convention, 1975.

Vaughan, R.W., Ed.D. "Semantic Confusion Involving Cost Analysis for Special Education Programming." Paper, 1975.

Walker, J. "Influence of alternative structural, organizational, and managerial options on the role of evaluation." A paper presented for the Symposium of SIG/Educational Research and Development Evaluators, Annual Meeting of the American Educational Research Association, Chicago, April, 1974.

Webster's New World Dictionary of the American Language: Second College Edition. New York: World Publishing Company, 1970.

Webster, W. J. "Technical auditing procedures." In Educational Product Audit Handbook, 38-103. Arlington, VA: Institute for the Development of Educational Auditing, 1971.

Webster, W.J. "The organization and functions of research and evaluation in large urban school districts." Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC, March-April, 1975.

Webster, W.J. and Cooley, W. Sensible school district evaluation. AERA Cassette Tape #EC2 (based on an Invited Address from the 1977 San Francisco Evaluation Conference) American Educational Research Association, Washington, DC.

Weiss, Carol H., Evaluation Research. Methods for Assessing Program Effectiveness, Prentice-Hall, Inc. Englewood, New Jersey, 1972.

Wiley, D. E. and Bock, R.D. "Quasi-experimentation in educational settings: Comment." The School Review, Winter, 1967, 353-366.

Wilson, J.Q. "On pettigrew and armour and afterwards." The Public Interest, 29, 1973, 132-134.

Wolf, R.L. "How teachers feel toward evaluation." In E. R. House (Ed.), School evaluation: The politics and process. Berkeley, CA: McCutcheon, 1973.

APPENDIX B

Glossary of Key Special Education and Evaluation Terms

GLOSSARY BY SECTION

Related Service Definitions 114

Federal Definitions of Handicapping Conditions. 118

P.L. 94-142 Program Areas 122

Definitions of Coding Levels on
Special Services Information System 127

Key Evaluation Terms. 128

RELATED SERVICES DEFINITIONS

AUDIOLOGY

Authority

P.L. 94-142 Regulations (§121a.13)

"Audiology" includes:

(i) Identification of children with hearing loss;

(ii) Determination of the range, nature, and degree of hearing loss, including referral for medical or other professional attention for the habilitation of hearing;

(iii)Provision of habilitative activities such as language habilitation, auditory training, speech reading (lip-reading), hearing evaluation, and speech conservation;

(iv) Creation and administration of programs for prevention of hearing loss;

(v) Counseling and guidance of pupils, parents, and teachers regarding hearing loss; and

(vi) Determination of the child's need for group and individual amplification, selecting and fitting an appropriate aid, and evaluating the effectiveness of amplification.

COUNSELING SERVICES

Authority

P.L. 94-142 Regulations (§121a.13)

"Counseling services" means services provided by qualified social workers, psychologists, guidance counselors, or other qualified personnel.

EARLY IDENTIFICATION

Authority

P.L. 94-142 Regulations (§121a.13)

"Early identification" means the implementation of a formal plan for identifying a disability as early as possible in a child's life.

MEDICAL SERVICES

Authority

P.L. 94-142 Regulations (§121a.13)

"Medical services" means services provided by a licensed physician to determine a child's medically related handicapping condition which results in the child's need for special education and related services.

OCCUPATIONAL THERAPY

Authority

P.L. 94-142 Regulations (§121a.13)

"Occupational therapy" includes:

(i) Improving, developing or restoring functions impaired or lost through illness, injury, or deprivation;

(ii) Improving ability to perform tasks for independent functioning when functions are impaired or lost; and

(iii)Preventing, through early intervention, initial or further impairment or loss of function.

PARENT COUNSELING AND TRAINING

Authority

P.L. 94-142 Regulations (§121a.13)

"Parent counseling and training" means assisting parents in understanding the special needs of their child and providing parents with information about child development.

PHYSICAL THERAPY

Authority

P.L. 94-142 Regulations (§121a.13)

"Physical therapy" means services provided by a qualified physical therapist.

PSYCHOLOGICAL SERVICES

Authority

P.L. 94-142 Regulations (§121a.13)

"Psychological services" include:

(i) Administering psychological and educational tests, and other assessment procedures;

(ii) Interpreting assessment results;

(iii)Obtaining, integrating, and interpreting information about child behavior and conditions relating to learning;

(iv) Consulting with other staff members in planning school programs to meet the special needs of children as indicated by psychological tests, interviews, and behavioral evaluations; and

(v) Planning and managing a program of psychological services, including psychological counseling for children and parents.

RECREATION

Authority

P.L. 94-142 Regulations (§121a.13)

"Recreation" includes:

(i) Assessment of leisure function;

(ii) Therepeutic recreation services;

(iii)Recreation programs in schools and community agencies; and

(iv) Leisure education.

SCHOOL HEALTH SERVICES

Authority

P.L. 94-142 Regulations (§121a.13)

"School health services" means services provided by a qualified school nurse or other qualified person.

SOCIAL WORK SERVICES IN SCHOOLS

Authority

P.L. 94-142 Regulations (§121a.13)

"Social work services in schools" include:

(i) Preparing a social or developmental history on a handicapped child;

(ii) Group and individual counseling with the child and family;

(iii)Working with those problems in a child's living situation (home, school, and community) that affect the child's adjustment in school; and

(iv) Mobilizing school and community resources to enable the child to receive maximum benefit from his or her educational program.

SPEECH PATHOLOGY

Authority

P.L. 94-142 Regulations (§121a.13)

"Speech pathology" includes:

(i) Identification of children with speech or language disorders;

(ii) Diagnosis and appraisal of specific speech or language disorders;

(iii)Referral for medical or other professional attention necessary for the habilitation of speech or language disorders;

(iv) Provisions of speech and language services for the habilitation or prevention of communicative disorders; and

(v) Counseling and guidance of parents, children, and teachers regarding speech and language disorders.

TRANSPORTATION

Authority

P.L. 94-142 Regulations (§121a.13)

"Transportation" includes:

(i) Travel to and from school and between schools,

(ii) Travel in and around school buildings, and

(iii)Specialized equipment (such as special or adapted buses, lifts, and ramps), if required to provide special transportation for a handicapped child.

FEDERAL DEFINITIONS OF HANDICAPPING CONDITIONS

1. "Handicapped children" refers to those children who have been determined through appropriate assessment as having temporary or long-term special educational needs arising from cognitive, emotional, or physical factors, or any combination of these. Their ability to meet general educational objectives is impaired to a degree whereby the services available to the general education program are inadequate in preparing them to achieve their educational potential.
2. "Deaf" means a hearing impairment which is so severe that the child is impaired in processing linguistic information through hearing, with or without amplification, which adversely affects educational performance.
3. "Deaf-blind" means concomitant hearing and visual impairments, the combination of which causes such severe communication and other developmental and educational problems that they cannot be accommodated in special education programs solely for deaf or blind children.
4. "Hard of hearing" means a hearing impairment, whether permanent or fluctuating, which adversely affects a child's educational performance but which is not included under the definition of "deaf" in this section.
5. "Mentally retarded" means significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behavior and manifested during the developmental period, which adversely affects a child's educational performance.

6. "Multi-handicapped" means concomitant impairments (such as mentally retarded-blind, mentally retarded-orthopedically impaired, etc.), the combination of which causes such severe educational problems that they cannot be accommodated in special education programs solely for one of the impairments. The term does not include deaf-blind children.

7. "Orthopedically impaired" means a severe orthopedic impairment which adversely affects a child's educational performance. The term includes impairments caused by congenital anomaly (e.g., clubfoot, absence of some member, etc.), impairments caused by disease (e.g., poliomyelitis, bone tuberculosis, etc.), and impairments from other causes (e.g., cerebral palsy, amputations, and fractures or burns which cause contractures).

8. "Other health impaired" means limited strength, vitality or alertness, due to chronic or acute health problems such as a heart condition, tuberculosis, rheumatic fever, nephritis, asthma, sickle cell anemia, hemophilia, epilepsy, lead poisoning, leukemia, or diabetes, which adversely affects a child's educational performance.

9. "Seriously emotionally disturbed" is defined as a condition exhibiting one or more of the following characteristics over a long period of time and to a marked degree, which adversely affects educational performance:

   a) An inability to learn which cannot be explained by intellectual, sensory, or health factors;

   b) An inability to build or maintain satisfactory interpersonal relationships with peers and teachers;

   c) Inappropriate types of behavior or feelings under normal circumstances;

d) A general pervasive mood of unhappiness or depression;

or

e) A tendency to develop physical symptoms or fears associated with personal or school problems.

f) The term includes children who are schizophrenic or autistic. The term does not include children who are socially maladjusted, unless it is determined that they are seriously emotionally disturbed.

10. "Specific learning disability" means a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations. The term includes such conditions as perceptual handicaps, minimal brain disfunction, dyslexia, and developmental aphasia the result of brain injury. The term does not include children who have learning problems which are primarily the result of visual, hearing, or motor handicaps, of mental retardation, of emotional disturbance, or of environmental, cultural or economic disadvantage.

11. "Speech impaired" means a communication disorder, such as stuttering, impaired articulation, a language impairment, or a voice impairment, which adversely affects a child's educational performance.

12. "Visually handicapped" means a visual impairment which, even with correction, adversely affects a child's educational performance. The term includes both partially seeing and blind children.

13. "Special education" means instruction provided at no cost to the parents, which is specially designed to meet the unique needs of a

handicapped child, including classroom instruction, instruction in physical education, home instruction, and instruction in hospitals and institutions. The term also includes vocational education, if it consists of specially designed instruction, to meet the unique needs of a handicapped child, at no cost to the parents.

P.L. 94-142 PROGRAM AREAS

CHILD IDENTIFICATION

Authority

P.L. 94-142 Regulations (Section 121a.220)

The provision of Child Identification is required in the Federal Regulation of P.L. 94-142. The mandate states that procedures must be established which insure that all children residing within the jurisdiction of the LEA who are handicapped, and who are in need of special education and related services, are identified, located and evaluated.

CONFIDENTIALITY

Authority

P.L. 94-142 Regulations (Sections 121a.560 - 121a.576) Annual Program Plan (Part VII-F)

The provision of Confidentiality in P.L. 94-142 requires the local education agency to implement procedures to insure that confidentiality shall be protected at collection, storage, disclosure and destruction stages. All policies and procedures delineated in the P.L. 94-142 Regulations and the Annual Program Plan must be followed.

FULL EDUCATIONAL OPPORTUNITIES

Authority

P.L. 94-142 Regulations (Section 121a.123)

The provision of Full Educational Opportunity is required in the Federal Regulations of P.L. 94-142. The mandate includes services, facilities and personnel. There is also a requirement for a full service goal and related timetable.

FACILITIES

Authority

P.L. 94-142 Regulations (Section 121a.223)

Each application must provide a description of the kind and number of facilities, personnel, and services necessary to meet the Full Educational Opportunity Goal.

PERSONNEL DEVELOPMENT

Authority

P.L. 94-142 Regulations (Section 121a.380 - 121a.383)

Each annual program plan must include a description of programs and procedures for the development and implementation or a comprehensive system of personnel development.

PARENT INVOLVEMENT

Authority

P.L. 94-142 (Section 121a.226)

Each application must include procedures to insure that, in meeting the Full Educational Opportunity Goal, the local education agency makes provision for participation of and consultation with parents or guardians of handicapped children.

LEAST RESTRICTIVE ENVIRONMENT

Authority

P.L. 94-142 Regulations (Section 121a.550)

Each State educational agency shall insure that each public agency establishes and implements procedures which insure that (1) handicapped children are educated with children who are not handicapped; and (2) handicapped children are removed from the regular education environment only when the nature or severity of the handicap is such that education in regular classes with the use of supplementary aids and services cannot be achieved satisfactorily.

PUBLIC CONTROL OF FUNDS

Authority

P.L. 94-142 Regulations (Section 121a.228)

Each application must provide assurance satisfactory to the State educational agency that control of funds provided under Part B of the Act and title to property acquired with those funds, is in a public agency for the uses and purposes under this part, and that a public agency administers the funds and property.

EXCESS COST

Authority

P.L. 94-142 Regulations (Sections 121a.182 - 121a.186)
Article 77 of the Annotated Code of Maryland (Section 106G - 106G-7)

The provision of Excess Cost requires that a local education agency may only use state special education funds and Part B funds for the excess cost of providing special education and related services for handicapped children.

NONSUPPLANTING

Authority

P.L. 94-142 Regulations (Section 121a.230)

Each LEA must provide assurance satisfactory to the the state educational agency that the LEA uses funds provided under Part B to supplement and increase the level of State and local funds expended for the education of handicapped children, and in no case to supplant those state and local funds.

COMPARABLE SERVICES

Authority

P.L. 94-142 Regulations (Section 121a.231)

Each LEA may not use funds under Part B to provide services to handicapped children unless the agency uses state and local funds to provide services to those children which are at least comparable to services provided to other handicapped children in that LEA.

INFORMATION - REPORTS

Authority

P.L. 94-142 Regulations (Sections 121a.232-121a.233)

Information - Reports: Each application (local application must provide that the local education agency furnishes information as may be necessary to enable the state education agency to perform its duties under this part, including information relating to the educational achievement of handicapped children participating in the local education agency's programs for handicapped children.

RECORDS

Authority

P.L. 94-142 Regulations (Section 121a.233)

Each application must provide that the local education agency keep such records, and affords access to those records, as the state education agency may find necessary to insure the correctness and verification of the information that the local education agency furnishes under Information-Reports.

PUBLIC PARTICIPATION

Authority

P.L. 94-142 Regulations (Sections 121a.226, 245, 560-573 and 500-508)

The provisions of Public Participation provide for each LEA to make all documents related to the local application available to parents and the general public and that all evaluations and reports required under Section 121a.232 are public information.

INDIVIDUAL EDUCATION PROGRAM

Authority

P.L. 94-142 Regulations (Section 121a.340 - 121a.346)

The term "individual education program" means a written statement for a handicapped child that is developed in a meeting by (1) a representative of the public agency other than the child's teacher (who is qualified to provide or supervise the provision of special education); (2) the child's teacher; (3) one or both of the child's parents or guardian; (4) the child, when appropriate; and (5) other individuals at the discretionof the parent or agency.

LOCAL POLICIES CONSISTENT WITH STATUTE

Authority

P.L. 94-142 Regulations (Section 121a.236)

Each application must provide assurance satisfactory to the State Education Agency that all policies and programs which the local education agency establishes and administers are consistent with Section 612(1)-(7) and Section 613(a) of the Education of Handicapped Children Act.

NONDISCRIMINATION AND EMPLOYMENT OF HANDICAPPED INDIVIDUALS

Authority

P.L. 94-142 Regulations (Section 121a.239)

Each application must include an assurance that the program assisted by Part B funds will not discriminate on the basis of handicap and will make positive efforts to employ and advance qualified handicapped individuals.

PROCEDURAL SAFEGUARDS

Authority

P.L. 94-142 Regulations (Section 121a.500, Subpart E)

Each state educational agency shall insure that each public agency establishes and implements procedural safeguards which meet the requirements of this section.

NONPUBLIC SCHOOLS

Authority

P.L. 94-142 Regulations (Section 121a.140 and Subpart D)

The provision of Nonpublic Schools is required in Subpart D of the Federal Regulations of P.L. 94-142. The mandate includes handicapped children who are or have been placed in or referred to a nonpublic school or facility.

VOCATIONAL EDUCATION

Authority

Section 613(a)(2) states that in order to qualify for assistance under Part B of the EHA-B, a State must "provide that programs and procedures will be established to assure that funds received by the State or any of its political subdivisions under any other Federal program, including Section 121 of the Elementary and Secondary Education Act of 1975 (20 U.S.C. 241c-2), Section 305(b)(80) of such Act (20 U.S.C. 884 a(b)(8) or its successor authority, and Section 122(a)(4)(B) of the Vocatinal Education Act of 1963 (20 U.S.C. 1262(a)(b)(B), under which there is specific authority for the provision of assistance for the education of handicapped children, will be utilized by the State, or any of its political subdivisions, only in a manner consistent with the goal of providing a free, appropriate public education for all handicapped children, except that nothing in this clause shall be construed to limit the specific requirement of the laws governing such Federal programs.

PRIORITIES

Authority

P.L. 94-142 Regulations (Section 121a.127)

Each annual program plan must include information which meet the requirements under Sections 121a.320-121a.324 of Subpart C. The State priorities must meet the timelines under 121a.122 and the State must make progress in meeting those timelines.

DEFINITIONS OF LEVELS CODED IN THE SPECIAL SERVICES INFORMATION SYSTEM

These definitions are provided to further clarify those listed on Page 18 of the SSIS Manual. The student is used as a reference point for calculating his level of service. The concept of level as defined by Maryland Bylaw includes three components, (1) whether the service is direct or indirect, (2) the number of hours in which the student receives special education and related services and (3) the location or learning environment where the special education and related service is provided. This information (using the student as the reference point) makes it possible to use SSIS data for providing the OSE report on Least Restrictive Environment and assists LEAs and the SEA in calculating the cost of special education by level.

Level I — Code Level I when the handicapped student is receiving only indirect service through the provision of consulant services to his/her regular classroom teacher.

Level II — Code Level II when the handicapped student is receiving up to 5 hours of direct special education and related service per week within the general education facility.

Level III — Code Level III when the handicapped student is receiving more than 5 and up to 15 hours of direct special education and related service per week within the general education facility.

Level IV — Code Level IV when the handicapped student is receiving more than 15 hours of direct special education and related service per week within the general education facility.

Level V — Code Level V when the handicapped student is receiving his entire educational program in a special school or a special wing of a school building.

Level VI — Code Level VI when the handicapped student is receiving a special education program in a residential setting requiring 24 hour care.

Level VII — Code Level VII when the handicapped student is placed for a short term on home or hospital teaching, as defined by Maryland Bylaw. Preschool handicapped children receiving services in home based or home and school based programs should be coded level VII.

KEY EVALUATION TERMS

Context evaluation - the provision of baseline information that delineates the environment of interest, describes desired and actual outcomes, identifies unmet needs and unused opportunities, and diagnoses problems that prevent needs from being met.

Input evaluation - the provision of information for determining optimum methods of resource utilization for accomplishing set goals.

Process evaluation - the provision of information for determining defects in procedural design or implementation during program implementation stages and for aiding in the interpretation of program evaluation data.

Product evaluation - the provision of information for determining the relative success of educational programs. Product evaluation may be interim or summative and generally addresses the extent to which programs meet their objectives, the relative effects of alternative programs, and/or program cost-effectiveness.

Applied research - the provision of information pertaining to interactions between student characteristics, teacher characteristics, and/or instructional strategies.

Basic research - the provision of information pertaining to fundamental relationships affecting student learning.

Data processing - the operation of the basic data retrieval and analysis system.

Planning services - the provision of technical assistance to district and project management in planning and managing projects and programs.

Ad hoc information - the provision of requested information on an ad hoc basis to decision-makers.

Research consultation - the provision of design and analysis help to district personnel in implementing their own research and evaluation projects.

Proposal development - the development of proposals, or the evaluation sections of proposals, for outside funds. In all cases outside funds should be a means to an end, not an end in themselves.

Cost-Benefit evaluation - the cost benefit analysis concerns both the expenditure and its benefit. The cost-benefit analysis attempts to identify the benefits of a program, both tangible and intangible.

APPENDIX C

Questions Raised by Local Directors of Special Education about Program Evaluation

ISSUES IN EVALUATION

The following questions were developed as a result of meetings with a key group of local special education administrators.

Q.: What should be evaluated?

Before beginning evaluation, it is important to have an adequate understanding of what the program or activity is that you want to evaluate. Describing and understanding what is to be evaluated may be deceptively easy to target; therefore, a statement should be written that describes why the program is needed, what it is supposed to accomplish, and how the program is to accomplish what is intended. A program goal can be broad and yet include specific objectives derived from it.

Q.: Are there general guidelines in planning for a successful evaluation?

There are several critical elements that can assist in planning your own evaluation or utilizing someone else's evaluation results. An evaluation should ask specific questions about a program. Always state the evaluation questions in advance. This will limit the scope of the evaluation and help everyone understand what kind of information they can expect to get and what kind they cannot expect to get. A most common error in evaluation planning is to ignore the context in which the administrator will make decisions. Therefore, in addition to the specific evaluation questions, we should ask how evaluation results will be used and what types of decisions will be made. One way to do this is by planning the report first.

Limit evaluation questions to those which will provide useful information for the people who expect to act upon it. The evaluator may help to formulate and clarify the evaluation questions, but the final choice of questions should be left to those who will use the data.

Every evaluation should ask questions about outcomes or products. Programs should be designed to achieve certain purposes or objectives, and an evaluation must determine the extent to which these have been met.

Evaluations of programs should ask questions about costs. It is important to know how much it cost to produce any observed results and whether the program can be reproduced in other places.

Q. What is the relation between program standards and evaluation?

Standards of program merit should be set for each evaluation. Standards are the criteria for making judgments about how well the program is working. They may be stated in terms of levels of success, comparisons among groups, and changes over time. Standards should be established before any data collection begins. When collecting data, it is easy to be influenced by what has been learned. If most participants liked the program, other important aspects of the evaluation may be overlooked. Another reason for setting standards in advance is to help select the evaluation design, sampling methods, instruments, and data analysis techniques that will produce the information needed to answer the questions and compare with the standards.

Evaluation standards must have scientific validity. Standards with scientific validity can be based on expert judgment, past experience, and empirical data. Empirical data can come from experimental designs in which one or more groups are systematically selected and observed over time, e.g., studies reported in the professional literature. Statistically significant differences over time or across groups are not synonymous with scientifically valid standards. Statistical findings are valid only when viewed in a systematic manner along with educational and program standards. A statistically significant outcome is one that is unlikely to occur by chance, and an educationally significant outcome is one that makes a meaningful difference in meeting program goals. Statistical significance is determined as part of the data analysis process, and it is not necessarily the same as programmatic significance. When there are a large number of people, it is possible to observe statistically significant outcomes that do not represent real gains in performance or any other changes that are truly important. See Appendix H for description of the National Program Evaluation Standards developed by NIE (National Institute of Education). These standards should always be referred to when reviewing an evaluation design.

Q.: How are evaluation standards set?

Standards are established to provide evidence of a program's success. A program is usually considered successful if it accomplishes its goal and the activities used are beneficial. Standards may encompass a particular view or a specific criteria by which to judge educational programs and projects. There are several ways to set standards: (1) utilize specialists; (2) set standards by

relying on past performance to determine reasonable expectations; (3) use standards which rely upon improvement measurements, and (4) rely upon established practice and norms. (See Appendix H)

Q.: How can a representative sample and its size be determined?

In evaluating relatively restricted programs or projects, use samples that reflect the composition of that specific population. The total population is required when the number of individuals in the program is too small in size.

Given the size of the effect considered meaningful (which should be stipulated by the standards) and an estimate of the standard deviation for the dependent variable in the population, it is possible to compute a sample size adequate to detect, with statistical precision, the sought-for changes. It is not always possible to obtain a sample large enough to be convincing. To make up for small samples, repeat observations several times with different (but equivalent) samples. Such repetition provides support for generalizing a program's results. Repetition with equivalent samples also increases the reliability of a program evaluation. (See Chapter V, Section V, on Selecting the Samples.)

Q.: How are evaluation instruments validated?

In developing an evaluation, many people decide to design their own evaluation instruments to suit their own particular purposes. It is important to validate these instruments through field testing and, if possible, expert review. You will want to know whether you are getting the information needed, whether the people using the instrument will be able to do so properly, and if the procedures are

standardized and consistent. Field testing should be carried out under similar conditions expected for the main study. A representative sub-sample of people who will participate in the evaluation study should be used. Prior to a field study test of your instruments, some experts should review the evaluation instruments and any accompanying instructions on how they are to be used.

Evaluation instruments must be both reliable (providing consistent measurements) and valid (providing accurate and relevant data). Three types of reliability are stability, equivalence, and homogeneity. Stability is examined by administering the instrument to the same group on two different occasions and correlating the scores, with a high correlation indicating reliability. Equivalency can be determined by giving two forms of an instrument to the same group on the same day or by giving two forms to two groups that have been randomly selected. The mean score and standard deviation on each form is compared, correlating the scores on each form with the scores on the others. If the forms have similar means and standard deviation, then they have equivalence reliability. Homogeneity relates to internal consistency, how well all the items on the instrument assess the same skill. To test for homogeneity, divide the instrument into two equal parts, and correlate the scores on one half with the scores on the other half. This procedure is called "split-half" reliability, and it estimates whether both halves of the instrument measure the same skills and characteristics. This formula is really the average score obtained from computing all possible split-half reliabilities.

Validity measures include predictive validity (predicting one's ability to perform a certain task), concurrent validity (comparing a new instrument against a known measure), content validity (proving

that an instrument's items accurately represent what they are intended to measure), and construct validity (establishing that the instrument measures information related to the specific construct). (See Chapter V, Section VI, for a more detailed discussion on selecting instruments.)

Q.: How can an appropriate evaluation design be selected?

The design strategy is the method by which you will group your sample and how variables will be manipulated to answer specific evaluation questions. Evaluators use internal and external validity as the criteria for deciding the degree of accuracy for an evaluation design which will answer the specific evaluation questions. Internal validity is the extent to which you can can distinguish between changes caused by the program you are evaluating and changes resulting from other causes. All designs must be internally valid.

External validity is the criterion for deciding whether an evaluation's findings will hold true for other people in other places. All designs must be externally valid. External validity is important whenever your findings are going to be applied to people or settings that were not included in your evaluation or when findings based on current participants will be used to make decisions affecting future participants.

Q.: How can an evaluation be managed effectively?

Evaluations are usually supposed to be completed within a given amount of time. It is important to know just when each evaluation activity will take place, the sequence of activities, and how long each activity will last. To do this, consider the following:

- the specific activities which will occur
- the deadline for completing each activity
- the amount of time to be given to each activity, i.e., the actual number of professional hours, days, or weeks needed, even if they are spread over a longer period of time.

Combine evaluation activities, deadlines, and time allocations in several ways to create schedules. List a beginning and ending date (or deadline) for each activity, and specify the time allocated to it. A time schedule can be organized by months. Determine the skills required to perform the various evaluation activities, and then assign staff and monitor the activities.

An evaluation usually has a budget allocation and must be completed without exceeding that amount. In preparing the budget, consider what needs to be done, activities, time allocations, and staff availability. (See Chapter V, Sections 1 and 2, on defining goals and objectives and agreeing upon evaluation standards.) The steps are summarized below:

Step 1: Did a change occur?

The first step is to prove that something changed as a result of the program. A change can be claimed for any outcome, product, or consequence that is logically related to the innovation.

To prove that a change occurred, you must have valid and reliable measures of conditions after the intervention has had time to take effect. You must also have estimates of what would have happened at the same point in time if there had been no intervention at all.

Step 2: Was the effect consistent enough and observed often enough to be statistically significant?

To be convincing, you either have to try educational interventions on large numbers of people or repeat them often enough with small samples to yield meaningful evidence.

Step 3: Was the effect educationally significant?

An educationally significant effect is one that is not trivial in magnitude, one that occurs in an important content area, and one that can be achieved at a reasonable cost.

Educational significance, unlike statistical significance, is not only a technical question but a matter of judgment. Educational significance is sometimes established on the basis of past experience, sometimes on expert review, or on theory or statistical rules-of-thumb.

Step 4: Can the intervention be implemented in another location with a reasonable expectation of comparable impact?

First, you must provide evidence that your project's setting is not unique. For example, if your high school is near a computer firm you might easily arrange a study program where students can learn keypunching and software design. The same program would be difficult to repeat, however, in a school without access to a computer company.

Second, you should be able to show that your staff is not unique. If only one teacher has used the program, it is hard to tell how much of a gain can be credited to the program and how much to good teaching. The same is true even if the teacher has repeated the program successfully several times.

Third, you want to show that the intervention can probably be duplicated. Many evaluations do not pay much attention to proving whether a program might be useful to others under different circumstances, even though some methods for establishing generalizability are available and would strengthen the evaluation design.

Step 5: How likely is it that the observed effects resulted from the intervention?

A number of explanations can be offered for any changes that appear to be associated with a program.

Alternative explanations are particularly important when you do not use rigorous experimental designs. Good evaluation design requires evidence that improved performance is a result of your particular educational intervention. You will have to rule out the possibility that other factors were actually responsible. Some of the factors that can be confused with program performance are maturation, statistical regression, mortality, instrumentation, testing, selection, the reactive effects of testing, the Hawthorne (Expectancy) Effect, and interference from other programs.

Step 6: Is the presented evidence believable and interpretable?

The sixth and last hurdle has to do with the general quality of the evaluation evidence.

- Present a complete description of the evaluation's methodology and findings.
- Describe instruments in detail. Document reliability and validity information, the number of students

tested and testing dates, and the level of forms of the tests used.

- Present data for all of the measures administered. Give types of scores (e.g., norms or raw scores). Clearly identify statistical tests, and provide sample sizes each time scores are reported.
- Be sure the inferences you draw from analyses are consistent with the evaluation's design and information.
- Be objective in your presentation, and always describe any limitations on the evaluation's design.

Q.: How are evaluation questions formulated?

The evaluation questions tend to start out being what you want to know. The following shows how this information can be transformed into questions related to how well goals were achieved and how successful the activities were.

Administrative type questions	Appropriate labels for this kind of evaluation	Kinds of questions the evaluator might pose
What needs attention? What should our program(s) try to accomplish? Where are we failing?	Input evaluation Needs assessment Organizational review Program documentation	What are the goals of the organization or community? How can goal priorities be determined? Is there agreement on the goals from all groups? To what extent are these goals being met? What are the areas in which the organization is most seriously failing to achieve goals? Where does it need to plan special programs or revise old programs? What are the program's goals and objectives? How are the program activities supposed to lead to attainment of the objectives?
What is happening in Program X? What does it look like? What do people experience? How much does it vary from site to site? How can program be improved?	Process evaluation Description of program implementation Formative evaluation	How many participants and staff are taking part? When? How often? Where? Is the program running on schedule? Is the program being implemented according to plan? Have the proper materials been purchased, and are they being used? What is a typical daily schedule? How are time, money, and personnel allocated in in the program? What activities do participants in the program become involved in? What are the program's most important characteristics--materials, activities, administrative arrangements? Are the program's important characteristics being implemented?

Administrative type questions	Appropriate labels for this kind of evaluation	Kinds of questions the evaluator might pose
What is happening in Program X, continued.		Are they leading to achievement of the objectives? What adjustments in the program might lead to the attainment of the objectives? Which activities are best for each objective? Are some better suited to certain participants? What problems are there, and how can they be solved?
Is Program X meeting its goals? Is Program X worth continuing or expanding? How effective is it? What conclusions can be made about Program X or its various components? What does Program X look like and accomplish?	Product evaluation Measurement of goal achievement Objectives-based evaluation Summative evaluation Outcome evaluation Consumer testing Evaluation research	What are the goals and objectives of Program X? How can they be measured? What do measures show about the degree of goal attainment? What are Program X's most important characteristics--materials, activities, administrative arrangements? Why should these particular activities reach its goals? Do the activities lead to goal achievement? What programs are available as alternatives to Program X? How effective is Program X in comparison with alternative programs?

Administrative type questions	Appropriate labels for this kind of evaluation	Kinds of questions the evaluator might pose
What is cost of implementing program?	Cost evaluation	How costly?
What is cost for specific type of service?	Cost analysis	What is direct cost for specific services by LEA?
What is cost per level of special education?	Cost efficient	What is indirect cost for specific service by LEA
	Cost effectiveness	What is cost of special education for specific type of handicap by level?
	Budget analysis	What is the cost of related service for specific type of handicap by level?
	Program audit	What is the cost per hour of a specific related service?
		What is the cost of a specific related service product? (product evaluation, etc.?)
		What is the cost of replicating a specific program at another site?
		What is the cost of a specific training package?
		Are funds being expended according to agreed upon budget allocation?

Q.: What should be included in the written evaluation report?

The evaluation report should be easily understood and communicate what was done, how it was accomplished, and why. The evaluation report must be credible and logically describe the evaluation process, to include:

- an introduction to the evaluation, the evaluation questions, and limitations on the scope of the evaluation
- the design strategy and the sampling procedures for each evaluation question and their limitations
- the information collection techniques and instruments and their limitations and any field activities
- the methods you used to analyze the evaluation information, their limitations, and the results for each analysis
- the answers to each evaluation question, including your interpretation of the findings and your recommendations

Q.: Should an external or internal evaluation be used?

In the case where an LEA does not have skilled evaluators available and funds are available, an external evaluator can be contracted with for all or part of the evaluation. An externally conducted evaluation generally has the advantages of being more independent, providing findings considered to be more credible by the evaluation audiences, and conserving key staff time. Some common limitations of externally conducted evaluations are that the evaluation personnel do not always understand the goals of the program or their rationale, it is often more expensive, external evaluators

may sometimes be viewed with suspicion if not distrust, the evaluation usually takes more time, and the findings may not be useful in leading to specific program modifications.

Q.: How should an external evaluation be selected?

When selecting external evaluation by individuals or agencies, consider the following factors:

- their training in evaluation methodology
- past experiences in evaluating similar programs
- skill in areas of special need, such as test development
- data analysis capabilities (if these are a relevant part of their task)
- skills in relating to students, staff, and other groups, such as parents or community volunteers
- availability and geographic proximity to the district
- their cost in relation to the services they agree to provide.

If a decision is reached to contract with an external evaluator, an LEA administrator should negotiate a written agreement delineating responsibilities of the LEA and of the external evaluator for the design, collection, analysis, and reporting of data.

Q.: What is a model?

Van Dalen (1966) defines models as simplified or familiar structures which are used to gain insights into phenomena that scientists want to explain. Models may be drawings or physical replicas that represent the real things, or they may be more abstract.

Mathematical equations, verbal statements, symbolic descriptions, graphic presentations, or electromechanical devices may be used to represent objects and relationships that are being modeled. Some investigators locate a structure about which much is known and use it to gain insight into a field about which little is known.

Well-known laws in the areas of physics, chemistry, biology, etc. are employed as models for constructing theories about education and psychology.

The term model is defined as "a structure that enables one to present concepts in such a way that researchers/evaluators can gain useful insight/information into their phenomena." The model includes:

1) structural elements (input, process, product),
2) program flows (from planning, through implementing to ending), and
3) criteria/standards for program evaluation.

Some characteristics of good models are that they describe and help us to understand complex systems and provide the framework within which experiments are done. Models help us to see new relationships as well as predict when experiments are impossible.

Some dangers in the use of models are that they invite over-generalization and entice us into committing logical fallacies. The relationships between variables may be incorrect or not validated. Model building often diverts useful energy into non-productive activity.

APPENDIX D

Cost Evaluation Techniques Used By

Maryland State Department of Education

COST EVALUATION TECHNIQUES

A. Direct Cost of Special Education

Special education direct costs are those costs which can be easily, conveniently, and accurately identified as part of the service areas and activities comprising the special educaton program. As outlined in the Financial Reporting Manual for Maryland Public Schools, special education direct costs are reported according to public, nonpublic, and state institution programs and further delineated by level of service. When recording costs, personnel salaries and wages are reported according to staff assignment, not caseload.

Within each level of service, all expenditures associated with it are recorded. They are classified as:

Salaries and Wages of teachers and aides on the payroll of the local education agency.

Contracted Services rendered by personnel not on the payroll of the local education agency.

Supplies and Materials purchased for and directly consumed in the special education program.

Equipment, including additional and replacement, bought for special education.

Other Charges for all expenses, other than those covered above, paid to or for the personnel involved in special education for the handicapped.

Transfer to Others

B. Indirect Costs of Special Education

Indirect costs of special education are those costs necessary for the provision of special education and related services, but are of such a nature that they cannot be readily or accurately identified with the specific service. Indirect costs of special education include a portion of:

Administration - Expenditures for the overall direction and control of the local education agency.

Nondistributed Instructional Expenditures - Expenditures for salaries of substitutes and other items that are consumed for instructional purposes but cannot be identified with any particular activity.

Guidance - Expenditures for activities of providing counseling to parents, students and other staff members on learning problems social development, career and course selection and the like.

Psychological Services - Expenditures for activities concerned with administering psychological tests and interpreting the results, planning programs to meet the special needs of pupils and the like.

Student Personnel Services - Expenditures for activities designed to improve pupil attendance including atendance services, social work services, and pupil accounting.

Health Services - Expenditures for activities that provide pupils with appropriate medical, dental, and nurse services.

Operation of Plant - Expenditures for care and upkeep of buildings and grounds, warehousing and distribution, and security and safety.

Maintenance of Plant - Expenditures of repair of buildings and equipment and replacement of equipment.

Fixed Charges - Expenditures for retirement funds, social security, and other employee benefits, insurance and judgements, rent, and other costs such as short term interest expense.

Generally speaking, the above indirect costs are allocated to special education in the same ratio that special education direct costs bear to total instructional direct costs. For example, if direct costs are $100,000, then 10% (10,000/$100,000) of the indirect costs would be allocated to special education as a cost of providing education to handicapped children.

C. Compatability of Special Education Financial Reporting with the Special Services Information System (SSIS).

As mentioned earlier, the direct costs of special education are reported, when appropriate, by level of service provider. The recording of costs in this manner corresponds to the recording of a handicapped child's level of placement in the SSIS, thus creating a method of estimating the cost per student for each level of placement.

D. Cost Evaluation Procedure

1. Direct Costs

The first step in doing a cost evaluation for a level of service, is to identify the direct costs associated with that particular level. This information is found in the LEA Annual Financial Report or the LEA Annual Budget depending on whether the prior year's or current year's information is being evaluated.

2. Support and Indirect Costs

Certain special education direct costs, by their nature, provide support to all facets of special education programs - namely Level I [Diagnostic/Prescriptive Services, Level II (Itinerant Services), Instructional Direction Services (central office staff)], and Improvement of Instructional Services (staff training and development, including workshops). Because of the overall support these areas provide to special education programs, it is necessary to prorate these costs over all levels of special education in order to get an accurate evaluation.

The portion of indirect costs allocated to special education as outlined in Part II must be distributed across the levels of service since this cost is incurred by all levels of service.

3. Methods of Allocation

There are numerous methods that can be used to allocate the above costs across the levels of service. No one method can be singled out as giving the most accurate results for all cost allocations. What is being evaluated and the nature of the cost being prorated should dictate which method should be used.

Suggested ways of allocating costs are by pupils served, by expenditures per level, or by services delivered:

a. Allocation by Pupils Served

This method distributes a cost based on the percentage of the handicapped population within each service level. For example, out of a total handicapped population of 1,000, 100 were being served in Level I. Ten percent (10%) (100/1,000) of the cost being distributed would be allocated to Level I. The end result of using this method is an equal cost being allocated to each student regardless of the amount of time the student spends in special education.

A refinement of this method would be to calculate the percentage of hours of special education delivered in each level of service to the total number of hours of special education delivered in all levels of service and distribute the cost to each level of service based on these percentages. This allows for costs to be distributed in proportion to the time spent in special education - the students spending the greater portion of their time in special education are generating the greater costs.

b. Allocation by Expenditures by Level

This method of allocating costs considers the direct expenditures charged to each level of service in relation to the total direct expenditures of special education. A percentage of direct cost charged to each level is calculated and this percentage is applied to the cost to be allocated. This method prorates in proportion to direct costs - the higher direct cost areas absorb more distributed cost.

c. Allocation by Services per Level

Basing a cost distribution on services provided assumes that each service generates the same amount of indirect/support cost regardless of the extent of the service. This can be done by calculating the percentage of special education services provided in each level and distributing the cost based on these percentages.

Whatever method(s) of allocation are choosen, they must be applied consistently throughout the analysis.

4. Pupils Served

The Special Services Information System (SSIS) is the source from which to gather information on handicapped students served. This information can be used for allocation purposes and to determine per capita costs by level. It is recommended that all SSIS reports used in a cost evaluation be taken at the same point in time in order to obtain accurate results.

E. Uses of the Cost Evaluation

Once the evaluation has been completed and the per capita cost for each level of service has been calculated, several purposes can be served. First, this information can be used to estimate the cost of educating children with a particular disability. For example, if the per pupil cost for each level of service has been calculated, by knowing the number of mentally retarded children being served by level, the estimated cost of educating mentally retarded children can be calculated.

The cost evaluation can be useful in estimating the cost to provide adequate special education services to unserved or inadequately served students. It can also be used to project the cost to provide services to a particular age group if the level of service required is known.

A cost evaluation is a valuable resource for current analysis and future projections. However, the evaluation is only as accurate as the data used in the computations. Even then, the end results are best considered as estimates.

APPENDIX E

Selected Annotated Bibliography

Adelman, Howard S. Evaluating Training Programs in Special Education. 1970. 14 pp. (ED 073 094).

The presentation encompasses a conceptualization of what is meant by evaluation and what is involved in evaluating training programs that prepare special educators. Specifically, evaluation is differentiated from research; Stake's general framework for evaluating educational programs is summarized; and a number of factors involved in conceptualizing the nature and scope of training program evaluation are considered, as are several critical problems related to such evaluation. Finally, some specific ideas are offered with reference to the evaluation of special education training programs. A 7-time bibliography is included.

Anderson, Richard D. and R. J. Sonnen. "Evaluation: Relating Special Education Programs to school Objectives," Journal of Special Education, 7 (summer 1973), pp. 197-203.

The article describes the fourth edition (1969) of Evaluative Criteria for the Evaluation of Secondary Schools (instrument of the National Study of Secondary School Evaluation) and notes that for the first time it includes a section on special education. An inservice training program for teachers is outlined in which teachers are introduced to Evaluative Criteria and provided means for adapting NSSSE's basic instrument to better evaluate and improve programs for the handicapped within their schools.

Blackhurst, Eric, W.: "Program Evaluation Through Learner Change: Considerations and Applications." Originally in Colorado Journal of Educational Research. 16,4,1977. Pp 1-5.

This article focuses upon product-evaluation, the outcome of the teaching process, pupil change. The Learner Referenced Model was designed and field-tested to evaluate effective teacher preparation programs at the University of Northern Colorado. Practicum students were given forms to complete with information on children with whom they were working in the public school settings. Faculty of the department of special education were asked to come to a consensus of objectives and to judge how well they believe teacher candidates would be able to perform there quantifiable standards. Completed forms were returned by the practicum students at the end of there public school experience and the frequency and percentage of behavior change for the individual objectives determined by the faculty were recorded as the students' forms were scored. The model facilitated program improvement in many areas of teacher preparation. Sample forms used and the designated objectives are included along with suggestions for the model's use in special education to collect data on growth in children--especially in new and innovative programs within the public school.

Borich, Gary D. The Appraisal of Teaching: Concepts and Process. Addison Wesley, 1977, 396 pp.

This volume is a sourcebook of conceptual frameworks and selected readings, intended to explore the issues and concepts introduced. Readings cover specific techniques, and readings that plan techniques into a framework of designing teacher-appraisal systems.

Educational Programs that Work. Far West Laboratory, Sixth Edition. Fall, 1979, Free.

Nearly 200 programs that measured up to the stringent criteria of the Joint Dissemination Review Panel. Department of Education, USDE.

Contact: Dr. Ray Hartjen
Maryland Facilitator Project
Educational Alternatives, Inc.
P.O. Box 265 ques of evaluation;' Chapter 3, data; and Chapter 4, current trends.

Grotsky, Jeffrey N. and others. Guidelines "PEP" Peer Evaluation Program. A Systematic Approach for Evaluating Educational Programs, 1973-1974. Harrisburg, Pennsylvania: State Department of Education, 1973. 40 pp. (ED 083 304)

The Peer Evaluation Program (PEP) has been instituted by the Division of Special Education, Pennsylvania State Department of Education, to allow intermediate units an opportunity to continuously improve their programs. General procedures of the program are outlined. Two evaluation forms are included: "Teacher Input for Self-Evaluation" and "Criteria for Use in Both Self and Peer Evaluations." A flowchart is provided.

Hauser, Carl. "Evaluating Mainstream Programs: Capitalizing on a Victory." The Journal of Special Education. Vol. 13, November 2, 1979. pp. 108-127.

This paper isolates eight factors and aspects of mainstreaming programs in a model for formative evaluation to provide program planning and on decision making for organizational charge. The complexity of the many interrelating components of mainstreaming is stressed with de-emphasis placed on general ability as a goal for evaluation in favor of focus on a programs specific interactions, their precise descriptions, and documentation. Gneral guidelines are presented for identifying purposes, sampling and instrumentation requirements. Each of the eight factors; Value, Idea, Circumstances, Timing, Obligation, Resistance, Yield and Ability are discussed descriptively and paralleled with aspects of mainstreaming programs.

in this book is that an evaluation must provide the people who commission it with information that is technically sound, useful and

believable. An Evaluation Primer Workbook: Practical Exercises for Educators. 1978, 57 pp. This companion book to An Evaluation Primer offers exercises intended to provide practice in applying evaluation skills to education programs.

Order From: Capitol Publications, Inc.
2430 Pennsylvania Avenue, N.W.
Washington, D.C. 20037
(202) 452-1000

Gallagher, James J., Richard C. Surles and Andrew E. Hayes. Program Planning and Evaluation. First Chance for Children, Volume 2. Chapel Hill, North Carolina: Technical Assistance Development System (625 W. Cameron Avenue, 27514), 1973. 52 pp.

This is the second in a series of monographs published for First Chance Network, the collective name for several model early education projects administered by the Office of Special Education in the U.S. Education Department. The monograph contains information presented in a series of meetings held for project personnel who had requested aid in the areas of planning and evaluation. It describes a model which is meant only to be an initial attempt to deal with these problems. Chapter 1 presents one method of problem-solving with special emphasis on planning educational programs that can be subjected to evaluation; Chapter 2, several techniques of evaluation; Chapter 3, data; and Chapter 4, current trends.

Grotsky, Jeffrey N. and others. Guidelines "PEP" Peer Evaluation Program. A Systematic Approach for Evaluating Educational Programs, 1973-1974. Harrisburg, Pennsylvania: State Department of Education, 1973. 40 pp. (ED 083 304)

The Peer Evaluation Program (PEP) has been instituted by the Division of Special Education, Pennsylvania State Department of Education, to allow intermediate units an opportunity to continuously improve their programs. General procedures of the program are outlined. Two evaluation forms are included: "Teacher Input for Self-Evaluation" and "Criteria for Use in Both Self and Peer Evaluations." A flowchart is provided.

Hauser, Carl. "Evaluating Mainstream Programs: Capitalizing on a Victory." The Journal of Special Education. Vol. 13, November 2, 1979. pp. 108-127.

This paper isolates eight factors and aspects of mainstreaming programs in a model for formative evaluation to provide program planning and on decision making for organizational change. The complexity of the many interrelating components of mainstreaming is stressed with de-emphasis placed on general ability as a goal for evaluation in favor of focus on a programs specific interactions, their precise descriptions, and documentation. General guidelines are presented for identifying purposes, sampling and instrumentation requirements. Each of the eight factors; Value, Idea, Circumstances, Timing, Obligation, Resistance, Yield and Ability are discussed descriptively and paralleled with aspects of mainstreaming programs.

How To: Evaluate Education Programs. A monthly guide to Methods and Ideas that work. Reference Service, $55.00.

Dr. Arlene Find and Dr. Jacqueline Kosecoff are editors of this newsletter which offers concise "how to" information on evaluation teaching, using questionnaires and interviews, planning evaluation and more. A bound paperback volume of all back issues is included.

Write:

Capitol Puplications
Suite G-12
2430 Pennsylvania Avenue, N.W.
Washington, D.C. 20037

Jones, Regionald L., Gottlieb, Jay, "Evaluating Mainstreaming Programs. Models, Caveats, Considerations, and Guidelines." Exceptional Children, May 1978. Pp. 588-600.

This article presents both the practical and theoretical issues pertinent to evaluation of mainstreaming programs including (a) critique of previous efficient studies, citing values as well as weaknesses in three procedures. (b) the many problems and issues involved in the evaluation of educational treatments - goals objectives, and valuables of instructional time, integration with non-handicapped and teacher attitudes of acceptance/non-acceptance.

Considerations related to appraising dependent measures are also discussed - (attitudes, achievement, cost) as well as the legal requirements for evaluation stated in P.L. 94-142. A useful checklist of guidelines for evaluation included to avoid mistakes early on the process, encouraging evaluation of mainstreaming programs not as an impossible task, but a challenging and valuable one.

Lilly, M. Stephen, Evaluating Individualized Education Programs: A Premier on Individualized Education Programs for the Handicapped.

This article recommends methods for collecting data on IEPs for monitoring (1) student progress toward short term instructional objectives (2) student progress toward long range goals (3) provisions of service delivery as agreed upon in the IEP. P.L. 94-142 evaluation requirements are cited as well as the need to go beyond the requirement. A chart of suggested data sources are provided for LEA and SEA use, as well as recommendations for data at the teacher level. A Data Based Instructional Model is outlined in five steps as well as a graphic data keeping procedure. Additional suggestions are made for evaluation of long range goals and service provision.

MacMillan, Donald L., and Semmel, Melvyn I. "Evaluation of Mainstreaming Program" Focus on Exceptional Children. September 1977.

Formative and summative evaluations that include all groups affected by mainstreaming --the regular class members, administrators, regular teachers as well as the special education students--are suggested as the needed model for evaluation of mainstreaming programs. Formative evaluation is approached using Stufflebeam's model (1971) using context evaluation to provide a rationale for determining objectives, both from an administrative and child-oriented perspective. Input evaluation determines how resources can be most efficiently displayed to achieve the goals of the educational program and assessing designs. Process evaluation enables ongoing modifications in procedures that are failing and product evaluation concerns the assessment and interpretation of the outcome during and at the end of the cycle. Summative evaluation instruments for both the regular and special students are discussed critically and recommendations are made.

Meierhenry, Wesley C. Planning for the Evaluation of Special Educational Programs: A Resource Guide. Lincoln, Nebraska: University of Nebraska, 1969. 52 pp. (Ed 040 555).

Developed along with a tape-slide package, the guide covers evaluation of special education programs. Robert McIntyre discusses evaluation for decision making; Victor Baldwin considers sources of help and how to use them; and Helmut Hoffman treats objectives as guidelines for action, data collection, and budget planning and evaluation. Wesley Meierhenry describes what and when to evaluate as well as reporting and dissemination; he also lists references on evaluation.

Meyer, Judy. The Development of Evaluative Criteria for Training the Handicapped: A Summary of Project Activities. Houston, Texas: University of Houston, Center for Human Resources, 1972. 35 pp. (ED 074 273).

This report summarizes the year's activities of a project designed to: (1) review pilot programs in vocational education for the handicapped in Texas, (2) develop guidelines for Texas Education Agency to use in evaluating pilot proposals for training the handicapped under provisions of the Vocational Education Amendments of 1968, and (3) develop guideines for school districts to follow in designing and evaluating programs for the handicapped.

_____. Self-Evaluation Guide for Local Districts for Vocational Education of Handicapped Students. Houston, Texas: University of Houston, Center for Human Resources, 1972. 39 pp. (ED 074 275).

Intended for use by Texas school administrators in developing and operating programs to better meet the vocational education needs of handicapped students, this guide first examines the need for a clearly developed program design with specific objectives as the foundation

on which program evaluation is based. Following this examination is a description of the following evaluation stages: (1) evaluation of the program's progress toward meeting the stated objectives, (2) evaluation of interlocking services which contribute to the desired outcome, and (3) evaluation of the product. This description also includes evaluative questions pertaining to: (1) aspects of the vocational rogram such as student selection and referral, class content and methodology, and teacher competence, (2) coordination between the vocational and special education departments and amount of parental involvement in the program, (3) administrative coordination of program facilities and type of organizational structure, and (4) product evaluation. A program evaluation checklist is appended.

National Study of School Evaluation

This organization has published evaluative materials for middle schools and junior high schools including the Middle School/Junior High School Evaluative Criteria and Secondary School Evaluative Criteria.

For additional information contact:

National Study of School Evaluation
2201 Wilson Boulevard
Arlington, VA 22201

Parsley, Marilyn J. The Development of a Planning and Evaluating Model for Programs Directed Toward the Exceptional Child in the School System. Doctoral dissertation. Denver, Colorado: University of Denver, 1971. 196 pp. (UM: order no. 71-20,100)

The planning and evaluating model developed in this study was designed for special education programs in the school system directed toward the exceptional child. It was based on the discrepancy evaluation model proposed by Malcolm Provus (1969). Part I of the model consists of a Likert-type scale developed and administered to seven groups in a pilot school-- special education teachers, special education adminstrators, ancillary staff, school administrators, regular teachers, principals, and parents. Part II of the model was concerned with the formulation of a tool directed toward the investigation and evaluation of the process and product. Part III is a description of the leadership behavior of the director of the special education program. Part IV consists of recommendations, alternatives, and modifications to be considered before using the various tools described.

Proger, Barton P. "Program Evaluation: The Model-Building Game," Journal of Learning Disabilities, 4(June-July 1971), pp. 292-306.

Several general models for program evaluation are described and criticized. The value of total program evaluation for making more scientifically-based decisions in programming is demonstrated, and agencies offering help in devising evaluation schemes for ongoing learning disabilities programs are recommended.--EJ

Provus, Malcolm. Discrepancy Evaluation: For Educational Program Improvement and Assessment. Berkeley, California: McCutchan Publishing Corporation (2526 Grove Street, 94704), 1971. 392 pp. ($8.95).

This book is a practitioner's guide to the evaluation of public school programs based on the concept that evaluation is the art of describing a discrepancy between expectation and performance. The Discrepancy Evaluation Model presented here is applied to school district programs, state agencies, and federal educational programs. Discrepancy evaluation, as described in this book, establishes that both case study method and the methodology of experimental design, as well as other rigorous techniques, are required to conduct meaningful evaluations. The author states that educational programs are complex undertakings that must be evaluated in stages relative to the development and stability of the program investigated. Moreover, he states that values, not statistics, are the source of evaluations. The basis for the values that both shape a school program and influence estimates of its success are explored. The author feels that use of this model by school administrators and othe educational practititoners should result in improvement of existing school programs, establishment of new and better programs, greater accountability of educators to the public, and wiser decisions by school administrators.

Prusso, Kenneth W. and others. Evaluation of Title I ESEA Project, 1971-1972. Volume III, Instructional Practices and Student Cognitive Performance. Philadelphia, Pennsylvania: Philadelphia School District, Office of Research and Evaluation, 1973. 152 pp. (ED 083 271)

ESEA Title I Projects carried out in Philadelphia during 1971-1972 are evaluated. The six projects are: Class for Mentally Retarded/Emotionally Disturbed Children; English as a Second Language; Improvement of Reading Skills (Reading Skills Centers; Shared Time, Part-Time, Primary Reading Skills Center): Individualized Education Center; Learning Dimensions; and Special Mathematics Project SEED. Related project reports included are: Computer-Assisted Instruction; English as a Second Language: Readiness; Reading Specialist Training and Development for Inner-City Teachers; Summer Reading Readiness; Summer Special Education; Teaching Basic Reading Skills--A Systems Approach; identification and discription of the project, current evaluation procedure, and conclusions.

Steward, Don. Instruction As A Humanizing Science. Slate, 1976, pp. 1,216.

Series of three volumes to be used to increase learning by increasing the effectiveness, efficiency, and productivity in designed instruction. Volume I, The Changing Role of the Educator: The Instructioneer; Volume II, A Behavioral Learning Systems Approach to Instruction: Analysis and Synthesis; Volume III, Creating an Emphasis on Learning: Quality Control, Productivity, and Accountability.

The complete series may be ordered from:

Slate Services Publishers
P.O. Box 8796
Fountain Valley, CA 92708

Talmage, Harriet. Statistics As A Tool For Educational Practitioners. McCutchan Publishing Corp. 1976, 244 pp.

This book addresses the fact that educators continuously engage in activities concerning evaluation that range from systematically planned research studies to impressionistic reactions. The reader is guided through a teaching strategy that engages the reader in a dialogue with the author as the reader is helped in making use of statistics as a resource for educational evaluation. Part I introduces ways of quantifying performance and products of performance; Part II looks at the correlation of multiple performance and the extension of the concept of correlation to evaluation instruments; and Part III applies the statistical concepts developed in Parts I and II.

The Evaluation Improvement Program is a non-technical inservice effort to train teachers, principals and agency managers in the techniques of education program evaluation. It offers self-instructional material, including a Program Evaluator's Guide, an evaluator's workbook and a trainers guide, and a catalog of evaluation training materials from the National Institute of Education.

Contact:

W. W. Walton
Evaluation Improvement Program
P.O. Box 2845
Educational Testing Service
Princeton, N.J. 08540

The Evaluation Network offers information or resources, training programs and materials, new developments in evaluation technology and theory, and practial applications. Membership includes subscrition to Evaluation News, a quarterly publication edited by Michael Scriven.

Write:

Evaluation Network
c/o Phi Delta Kappa
Eighth St. and Union Ave.
P.O. Box 789
Bloomington, Indiana 47401

The Evaluation of Instructional Programs. Annotated Bibliography, Educational Research Service, Inc. 1974, $5.00.

A comprehensive bibliography containing descriptive annotations on more than 250 documents with several specific to special education considerations.

Contact:

Education Research Services, Inc.
1815 North Fort Myer Drive
Arlington, Virginia 22209

Turner, Ruth M., Macy, Daniel, J., A Five Year Longitudinal Study of the IEP Implementation. Paper. February, 1978, 93p. Ed. 170 983.

This paper describes a process evaluation built into a Texas school district's implementation of an individualized mainstream special education program. Five phases of implementation took place, and were evaluated concurrently with techniques specific to the sites and phase of the plan. Modifications were often made as results of the evaluations, including the establishment of an advisory council, materials center, and eventually a record keeping system for IEPs with an Act of sample objectives for 30 categories. Product Evaluation invalid review of the effectiveness and quality of th IEP's themselves resulting n an increase of staff to develop more IEP's and four format options for IEP's to increase their accuracy and effectiveness. Guidelines are given for IEP implementation in large urban settings.

White, Owen R. Curriculum-References, Time-Series Programs Evaluation. Paper. March, 1978. Ed. 167 620

This article cites the weaknesses of the use of control groups, regression-discent inquiry projection models, and norm-referenced testing in educational program evaluation, particularly with the severely-profoundly handicapped. An altenative of curriculum-referenced evaluations is offered based on compilation of a test of objectives to be checked off as they are mastered since the performances of all children are held relative to the same set of objectives, group summaries and analysis are also possible. The sequential nature of a curriculum becomes the common metric timeline for multiple-baseline studies, and since all pupils use the same curriculum the needs of all students can be met in all areas of concern, while still evaluating a program's overall effectiveness. Othr benefits include immediate feedback for educators on success of program goals, improved development of IEP's, curriculum refinement, and more longitudiant time series date studies.

APPENDIX F

A Description of the Material Included in the Program Evaluation Resource Kit

Bennett, Ann A. and Fred L. Fiter. Getting It All Together: The JDRP Process. Capla Associates, Inc. 44 pp. $2.00

A guide for the utilization of the submission process to the USOE/NIE Joint Dissemination Review Panel for review and approval of exemplary education programs to be included in the USOE National Diffusion Network

Distributed by:

Central Service Unit
Capla Associates, Inc.
18 Overlook Avenue
Rockville Park, NJ 07662

Borich, Gary D. "Program Evaluation: New Concepts, New Methods." Focus on Exceptional Children, Vol. 9, No. 3, May, 1977. pp 14.

The ways to decompose a program by looking at each component within the hierarchical structure are explored. This methodology can be used by planners, developers, and evaluators for thinking in a structured way about large and complex programs through providing strategic concepts for assessing the results of the planning, designing, and development process.

Bowers, John J. Planning a Program Evaluation: An Educator's Handbook. Research for Better Schools, Inc. March, 1978, 125 pp.

A handbook designed to help school districts plan their curriculum program evaluations. There are four major sections: Case Study - Narrative; Case Study - Examples; Instructions for Completing Forms; Seeking Further Assistance.

Dunst, Carl J. "Program Evaluation and the Education for All Handicapped Children Act." Exceptional Children September, 1979. pp. 24-31.

This article describes the evaluation requirements of P.L. 94-142 and is specific to the type of evaluation that is necessary to meet requirements of the Act. Specific examples of strategies to measure program effectiveness are given. Evaluation is defined and major types of evaluation identified.

Elliott W. Eisner. "The Use of Qualitative Forms of Evaluation For Improving Educational Practice." Educational Evaluation and Policy Analysis. Nov.-Dec., Vol. 1, No. 6, 1979. pp. 11-19.

Covers four major points with regard to educational evaluation: (1) the view of knowledge as existing only where logical criteria and measurement exists as a constraint to educational research; (2) sources of knowledge are at least as diverse as the range of information provided by the senses; (3) concepts are not confined to linquistic or numerical terms; (4) the vocabulary and syntax of a symbol system is required to make public what has been conceptualized.

Evaluation Comment, The Journal of Education Evaluation, published by the Center for the Study of Evaluation, University of California, Los Angeles. CSF is the only federally funded research and development Center working exclusively on problems in educational evaluation.

Please write to:

James Burry, Managing Editor
Evaluation Comment
Center for the Study of Evaluation
145 Roof & Hall
University of California, Los Angeles
Los Angeles, California 90024

Holowenzak, Stephen P. A Strategy For the Evaluation of Instructional Programs. Report. Maryland State Department of Education, January 1979-79-4, 4 pp.

This report includes a flowchart of program evaluation activities for the Division of Instruction. An evaluation profile plan is also included.

House, Ernest R. Assumptions Underlying Evaluation Models. Educational Researcher, March 1978.

The author selects major evaluation models and summarizes each: systems analysis; behavioral objectives; decision making, goal free, art criticism, accreditation, adversary, and transaction. The interrelatedness of models is shown in a taxonomy, and each related to the philosophy of liberalism.

Howe, Clifford, E. and Marigail E. Fitzgerald. "Evaluating Special Education Programs." Focus on Exceptional Children Vol. 8, No. 9, February, 1977.

This article presents a multi-level evaluation model. The model is implemented in Iowa in 15 intermediate units, and addresses two major issues specifically, that of determining the technical approach which appears to have the highest likelihood of yielding useful data to use in making decisions regarding future directions for the program; and developing a readiness in the organization to undertake program evaluation.

Morocco, Catherine Cobb. The Role of Formative Evaluation in Developing and Assessing Educational Programs. Curriculum Inquiry Vol. 9, No. 2, 1979.

This article offers a development view of program evaluation, which suggests that each stage in the program life cycle requires a different research purpose and methods, and that the real effectiveness of program research lies in the continuity between

research carried out in early and later stages of program development. The model the author offers integrates the conflicting aims and methods of program research into a description of how research feeds into both the development and the final assessment of the program.

Olsen, Kenneth R. Evaluating Effectiveness of Special Education Programs. Mid-South Bulletin No. 10, Mid-South Regional Resource Center.

An excellent clarification of the relationships between special education program monitoring and evaluation of effectiveness as required by P.L. 94-142. A structure for total program evaluation which incorporated both program monitoring and evaluation effectiveness is proposed, and classic evaluation models as well as recent developments in the field are incorporated.

Program Evaluation Kit

Editor: Lynn Lyons Morris
Center for the Study of Evaluation
University of California at Los Angeles.

The Program Evaluation Kit contains eight books written to guide and assist practitioners in planning and managing evauations. The Kit is designed to answer the hundreds of conceivable questions that a practicing evaluator might ask. Its practical, field-tested step-by-step guides offer detailed advice, supported by clear definitions and useful procedures explained in non-technical language. Each of the Kit's books can be used independently, and all have proven valuable to practitioners with varying levels of experience. They have served, as well, as basic texts for evaluation methods courses.

Volumes may be purchased individually-or the complete Program Evaluation Kit may be purchased as a unit at $49.95. The Kit consists of:

1. Evaluator's Handbook
2. How to Deal with Goals and Objectives
3. How to Design a Progam Evaluation
4. How to Measure Program Implementation
5. How to Measure Attitudes
6. How to Measure Achievement
7. How to Calculate Statistics
8. How to Present an Evaluation Report

Order From:

Sage Publication
275 South Beverly Drive
Beverly Hills, CA 90212

Smith, Judy, Claudia Arkell, and Carter Allen. "The Educational Developer As Change Agent." Paper. Exceptional Children (In Press). 1980, 24 pp.

Educational development is reviewed and explored as "the process of creating new alternatives that contribute to the quality of education." It is indicated that responsible dissemination i.e., telling, showing, helping, training, involving and intervening. Demonstratable effects of educational practices are thoroughly discussed.

Salvia, John and James E. Ysseldyke. Assessment in Special and Remedial Education. Houghton Mifflin, 1978, 531 pp.

A text intended as a first course in assessment for those whose careers require understanding and informed use of assessment data. The primary audience is teachers in special and remedial educaton: the secondary audience is the support system for special needs student. Part 3 provides detailed discussions of various assessments and Part 4 deals with the application of assessment practices in special and remedial education.

Tallmadge, G. Kasten. Ideabook. The Joint Dissemination Review Panel. U.S. Department of Health, Education and Welfare. October, 1977, 113 pp.

This publication of the National Institute of Education, U.S. Office of Education, illustrates the ways to gather evidence of the effectiveness of educational innovations, and bring existing evidence together in a dynamic way. Using the explicit guidelines and model programs can lead to program adoption by the Joint Dissemination Review Panel.

Williams, Martha K. Educational Manager's Guide To Project Evaluation. 1978, 29 pp. $2.00.

This concise handbook gives a clear overview of the evaluation process, a chart of information needs identification, evaluation design and planning format, and post evaluation considerations. Ten tips for consumers of evaluation consultant services are also given.

Write:

Dr. David P. Crandall
Executive Director
The Network
290 S. Main Street
Andover, MA 01810

APPENDIX G

Overview of the Status of the

Monitoring and Evaluation Procedures

of the Division of Special Education

Maryland State Department of Education

Monitoring and Evaluation of Special Education Programs at the State Level

The Maryland Bylaw requires each LEA to develop an annual comprehensive plan. A program planning and budgeting system is used in this plan to facilitate comparing information between LEAs and interfacing with other agencies at the state and federal levels.

MSDE conducts a statewide needs assessment on personnel training. The needs assessment survey uses a stratified sampling design according to the type of staff and LEA. MSDE then generates reports to assist each LEA in identifying training needs to be incorporated into their comprehensive plans.

MSDE also provides each LEA with comprehensive reports of the number of students and types of services provided, including handicapping conditions and learning environments, through its SSIS System for incorporation into LEA comprehensive plans. The SSIS System is a discrepancy model which allows analysis of the difference between services recommended and services provided.

P.L. 94-142 requires that the State Education Agency (SEA) evaluate the effectiveness of local programs in meeting the needs of handicapped children. The regulations state:

Each state educational agency shall:

1. Undertake monitoring and evaluation activities to insure compliance of all public agencies within the State with the requirements of Subparts C, D, and E.

2. Develop procedures (including specific timelines) for monitoring and evaluating public agencies involved in the education of handicapped children. These procedures must include:

    a. collection of data and reports;
    b. conduct of onsite visits;
    c. audit of Federal fund utilization; and
    d. comparison of a sampling of individualized education programs with the programs actually provided. (121a.601)

In addition, Maryland law requires the SEA to monitor all special education programs. These requirements are found in Bylaw 13.04.01, "Programs for Handicapped Children," which states:

> The State Superintendent of Schools shall be responsible for the development of policies and procedures for monitoring and evaluation activities which insure compliance of all public agencies within the state with the requirements of these regulations and State and Federal Legislation. These procedures shall also provide for the evaluation of the effectiveness of the individualized education programs. (13.04.01.03E)

In response to these mandates, a systematic and comprehensive monitoring and evaluation system has been developed by the Maryland State Department of Education (MSDE). The state helps the LEAs draw up its annual plan, and monitors its implementation through a site visit every three years. The purposes of the monitoring and evaluation site visit are:

1. To determine the degree of local and state compliance with the policies, standards, and guidelines set forth in the local and state annual program plans,

2. To determine the areas of either recommended or required change in the programs of special education and related services provided to handicapped children within the 24 local education agencies (LEAs), and to provide a basis for both technical assistance and personnel development,

3. To provide performance data to be used in the the ongoing planning and quality improvement of special education programs in the 24 LEAs, and

4. To determine the service delivery systems which provide the most effective services at the least cost.

MSDE has developed a Monitoring and Evaluation Manual to be used in conjunction with LEA site visits to monitor implementation of comprehensive plans.

Section I, Procedural Overview, provides the reader with a brief description of the components of the monitoring and evaluation instrument, a description of LEA and state presite and onsite responsibilities for the site visit, and a description of the procedures used to assemble the Report of the Findings;

Section II, Directions for Completing the Review Components, provides the reader with directions for completing the four components and includes summary data forms from the various components that are used by the chairperson of the Monitoring and Evaluation Site Visit Team to write the Report of the Findings.

In addition to Sections I and II, an appendix is provided which includes the data acquisition forms used onsite by the site team to collect the information for each of the four components.

Procedural Overview

Monitoring and evaluation activities that take place are designed to document the extent of compliance in each LEA with all federal and state regulations appropriate to the provision of programs of special education and related services to handicapped children. The monitoring and evaluation process involves those presite and onsite activities necessary to complete the four components of the monitoring and evaluation instrument. These four components are:

The Administrative Policy and Procedure Review
The Individual Case Review
The Parent, Administrator and Staff Questionnaires
The Federal Funding Review

The information gathered in these four components serves as the basis for the Report of the Findings. This report is provided to each LEA approximately 60 days after the completion of the site visit. The report shows the findings for all areas within each of the four components as detailed below, and specifically cites those areas in which a LEA has been found to perform in a commendable or exemplary fashion, as well as those areas in which change is either recommended or required.

Monitoring and Evaluation Components

The Administrative Policy and Procedure Review (APPR)

This component of the Monitoring and Evaluation instrument is used to determine the extent to which a LEA has in existence those policies, procedures, methods and activities it has specified in its comprehensive plan and the extent to which these are implemented and are consistent with both federal and state requirements for the provision of programs of special education and related services to handicapped children. The APPR focuses on the 21 program areas specified in the P.L. 94-142 regulations under LOCAL EDUCATIONAL AGENCY APPLICATIONS - CONTENTS. These program areas are generally descriptive of the federal regulations an LEA must comply with in meeting its responsibilities for providing special education programs. Additional requirements necessitated by Maryland Bylaw 13.04.01 have been included in the APPR under the appropriate program areas. The 21 program areas are found within Part E, the Federal Requirements section of the local plan, and they are also listed in the Index of Elements table at the rear of this appendix.

Individual Case Review (ICR)

This component of the monitoring and evaluation instrument is used to determine whether or not various activities and processes specified in P.L. 94-142 and Bylaw 13.04.01 relevant to delivery of special education and related services are first, carried out, and second, provided in the proper sequence and within the specified time limits. Cases assessed using the ICR are selected through the Special Services Information System (SSIS), using a stratified random sampling procedure that takes into account the population size of a particular LEA as well as all levels of special education services specified in the Bylaw.

Thirty-two separate items and relevant dates are assessed by the ICR. These items cover the span of events from the time a child is initially referred for assessment, through the development and implementation of the Individualized Education Program (IEP) and the annual review by the Admission, Review and Dismissal (ARD) Committee. Individual items in the ICR are scored on the basis of supporting documentation within each case folder. The cases of children receiving special education services prior to, as well as after, the adoption of Bylaw 13.04.01 can be accommodated in the ICR.

Parent, Administrator, and Staff (PAS) Questionnaires

This component consists of three questionnaires that are distributed respectively to a random sample of parents of children receiving special

education services, administrators of programs providing special education services, and staff members providing direct service to handicapped children in special and general education programs.

It is understood that the policies and procedures developed by any LEA are useful to the degree that they are disseminated, understood and adopted within a school system and to the degree that they ultimately bear on the programs and services delivered to any particular child. Whereas the APPR generally assesses the presence of various policies, procedures and resulting methods and activities specified in a LEA's Local Comprehensive Plan, the PAS Questionnaires, in conjunction with the ICR, assess the results of these policies, procedures, methods and activities.

Federal Funding Review (FFR)

Federal law requires every SEA to monitor and evaluate the use of federal dollars distributed to states for the provision of special education services. Each educational unit (city/county/institution) must be site visited for the purpose of evaluation not less than once every three years in order to determine local compliance with the laws and regulations governing the use of P.L. 94-142 and P.L. 89-313 funds. Further, a determination is simultaneously made regarding local observance of relevant state education law, bylaws and adopted resolutions by the State Board of Education. To this end, the Federal Funding Review consists of six parts: Parts I and II, General Funding and Allocation; Part III and IV, P.L. 94-142 Review; and Part V and VI, the P.L. 89-313 Review.

Issues of achieving compliance, meeting local needs and the accomplishment of objectives are measured against the descriptions specified in each local comprehensive plan and accompanying projects (Pass Through and Discretionary - when applicable) for P.L. 94-142 and any projects funded through the use of P.L. 89-313 funds.

The Review Process

This process encompasses both LEA and MSDE responsibilities in the preparation and review of all components in the monitoring and evaluation instrument. This process culminates in the Report of the Findings that the MSDE prepares and submits to the LEA.

LEA Responsibilities

LEA responsibilities vary with each of the four components of the monitoring and evaluation instrument and include both presite and onsite preparation.

LEA Presite Preparation

At a meeting arranged by MSDE at least six weeks before the scheduled site visit, the LEA will receive a packet containing:

The APPR Data Acquisition Forms
The PAS Questionnaires
The FFR Data Acquisition Forms

These forms are to be handled as follows:

The APPR Data Acquisition Forms

The APPR Data Acquisition Forms contain objectives and a number of possible methods and activities for each of the 21 program areas. All objectives in the APPR must be addressed by the LEA. The methods and activities listed on the APPR Data Acquisition Forms are suggested. Space has been provided so that methods and activities not printed on the forms may be added as necessary. With the assistance of the state regional administrator, the LEA should indicate on these forms which methods and activities it has written into its current plan that it will accomplish during the school year, which methods and activities were recommended by the state as a result of the most recent monitoring and evaluation site visit, and which methods and activities it has engaged in that supplement to those submitted. The LEA will then, with the assistance of its Regional Administrator, furnish objective measures to indicate that it has engaged in the methods and activities specified in the plan and indicate these in the appropriate column on the APPR Data Acquisition Form. This information should be returned to the state education agency at least two weeks prior to the scheduled site visit.

The Parent, Administrator and Staff Questionnaires

The PAS Questionnaires are to be distributed by the LEA to parents, administrators, and staff members. The following procedures are used in selecting recipients of questionnaires:

Parent Questionnaire

Using SSIS, MSDE will randomly select a sample of parents of handicapped students in each LEA. The criteria for determining the number of parents to be sampled will be at the .05 level of confidence.

The MSDE will give a list of students Soundex numbers and corresponding school name and teachers' initials with an equal amount of questionnaire packets to the LEA Director of Special Education. Each questionnaire will have a county, school and Soundex number affixed.

The LEA Director of Special Education will disseminate the questionnaires to the parents/guardians of the children according to the affixed Soundex numbers.

Administrator Questionnaire

Using SSIS, MSDE will randomly select a sample of school administrators serving handicapped students in each LEA. Criteria for determining the number of administrators to be sampled will be at the .05 level of confidence. Each questionnaire will have a county and school number affixed.

The LEA Director of Special Education will disseminate the questionnaires to the administrators identified.

Staff Questionnaire

Using SSIS, MSDE will select a stratified random sample of speech therapists, special, and regular education teachers providing services to handicapped students in each LEA. The criteria for determining the number of special and regular education teachers to be sampled will be at the .05 level of confidence.

The MSDE will send a list of schools with an ample supply of questionnaire packets to the LEA Director of Special Education. A portion of the questionnaires will have county, position, staff initials of the special educators and related service personnel and school affixed. The remainder will be used as described in the following step.

The building principal will randomly select the regular education faculty working with handicapped children, and will disseminate the number of questionnaires selected by MSDE and provided through the LEA Director of Special Education.

The parents/guardians, administrators, and school staff selected are to complete the questionnaires and return them to the Division of Special Education in the stamped envelopes provided. This information should be received by MSDE no later than four weeks after the questionnaires are given to the LEA.

If an insufficient number of questionnaires is returned, the MSDE will issue a duplicate set of questionnaires for non-respondents to be processed in the manner outlined above.

Data Acquisition Forms

The Federal Funding Review (FFR) will be sent to each county/city/institution five weeks prior to the scheduled visit. The recipient will complete the forms as described below and return them to the Division of Special Education not later than two weeks prior to the scheduled visit.

Parts I - IV, which relate to general funding allocation and P.L. 94-142 Use of Funds, will be sent to each LEA. Parts V - VI, which relate to P.L. 89-313, will only be forwarded to those LEAs eligible to receive 89-313 funds.

Report of the Findings

The Report of the Findings is designed to resolve various findings of all components of the monitoring and evaluation instrument and to focus these on the 21 federally required program areas that are the basis for the Administrative Policy and Procedure Review and the local plan. To accomplish this resolution, all elements of each of the four components of the monitoring and evaluation instrument are referenced to one or more of the 21 program areas. This enables the monitoring and evaluation team to

draw from all sources, judging, first, whether the policies, procedures, methods, and activities specified in the local plan have been met and, second, whether in meeting these, a LEA has brought itself into greater or lesser conformance with state and federal regulations governing services to handicapped children. A LEA will have in the Report of the Findings accurate and consistent information that it can use in preparing and revising its plan for the coming year.

In preparing the Report of the Findings, the team chairperson will call on all data collected by the monitoring and evaluation instrument as well as any other data that have been provided by the LEA or collected by the state prior to and during the site visit. Each LEA has many elements in its system of service delivery that it can be justifiably proud of. The Report of the Findings addresses those areas for which a LEA is to be particularly commended and those areas for which change is recommended or required. The chairperson may, in the final report, cite any of the four components of the monitoring and evaluation instrument. In every case, the data summary from each of the four monitoring and evaluation components will be attached to the text of the report.

A LEA can expect to receive the Report of the Findings approximately 60 days after the monitoring and evaluation site visit. The LEA will then have 30 days to initially respond to the report. The LEA responses must, at minimum, address those areas of the report in which change is required. If, in the state's opinion, it is determined that the areas of required change have not been appropriately addressed, further responses will be required.

The following Index of Elements shows the relationship of each element within the monitoring and evaluation instrument to one or more of the 21 program areas. Each program area has a list of activities and objective measures that the LEA is using as standards. The Index of Elements shows types of data gathering methods used by MSDE which identify data source and program objectives being evaluated.